Semantic Parsing

ANLP

Horacio Rodríguez

Outline

- Introduction
- Approaches to Semantics
- Semantic spaces
- Semantic Role Labeling
- Meaning representation
- Semantic Parsing
- Learning Semantic Parsers
- Embedded systems, Deep Learning
- NL accessing to Linked Data

Introduction

- Obtaining and Representing the meaning of a sentence
  - Meaning Representation
  - Semantic Interpretation
- Desideratum
  - Rich meaning representation: FOL
  - Unrestricted texts
  - Full Semantic Parsing
- But ...
  - Less expressive formalisms: DRT, DL
  - Domain restricted
  - Shallow approaches
  - Intermediate processes:
    - Lexical Semantic Tagging
    - Semantic Role Labeling

Semantic Grammars

- Combination of syntax and semantics in a unique formalism
- Terminal symbols are semantic tags.
- Robust systems in restricted domains
- Easier to build the meaning representation

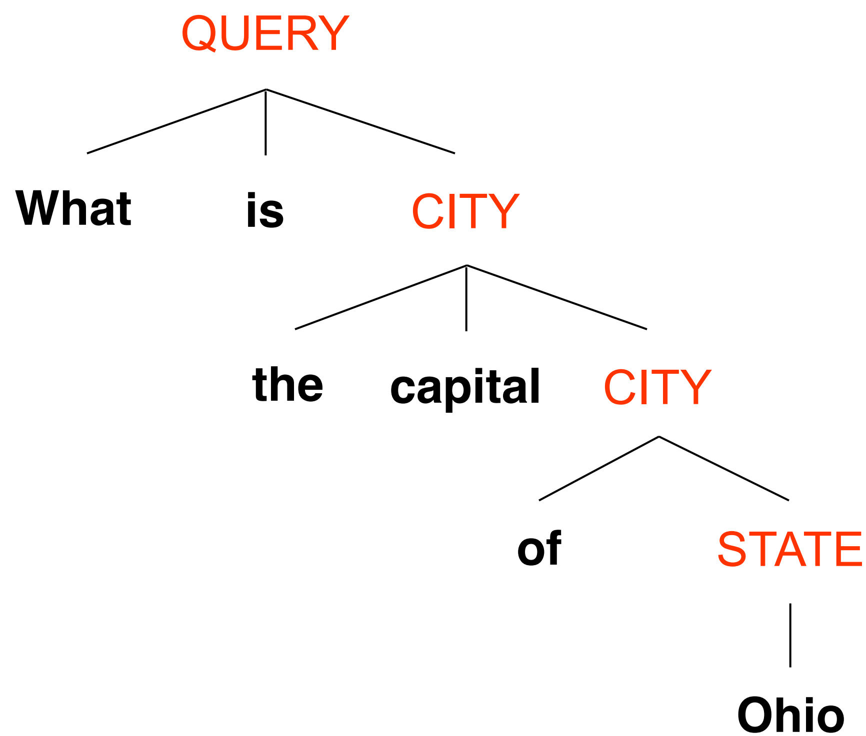
Context-Free Semantic Grammar

QUERY → **What is** CITY

CITY → **the capital** CITY

CITY → **of** STATE

STATE → **Ohio**



Approaches to Semantics

- Compositional Semantics
- Distributional Semantics

Approaches to Semantics

- Compositional Semantics
  - Semantic complex entities can be built from its simpler constituents
    - Ted Briscoe (2011) Introduction to Formal Semantics for Natural Language
    - Gennaro Chierchia and Sally McConnell-Ginet. (2000) Meaning and Grammar - an Introduction to Semantics (second edition). The MIT Press, 2000.

Approaches to Semantics

– Distributional Semantics
  - Distributional Hypothesis: the meaning of a word can be obtained from the company it has
  - M. Baroni and A. Lenci. 2010. Distributional Memory: A general framework for corpus-based semantics. Computational Linguistics 36 (4): 673-721.
  - M. Baroni and R. Zamparelli. 2010. Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2010), East Stroudsburg PA: ACL, 1183-1193
  - William Blacoe and Mirella Lapata. 2012. A Comparison of Vector-based Representations for Semantic Composition. In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, 546--556. Jeju Island, Korea.

Approaches to Semantics

- Distributional Semantics
  - These models are most commonly used for individual words and short phrases, where vectors are created using distributional information from a corpus.
  - While vector space representations for individual words are well-understood, there remains much uncertainty about how to compose vector space representations for phrases out of their component words.
  - Should all syntactic categories of words be represented as vectors, or are some categories, such as adjectives, different?
  - does semantic composition factorize according to a constituency parse tree?
  - See
    - Jayant Krishnamurthy, Tom M. Mitchell (2013)Vector Space Semantic Parsing: A Framework for Compositional Vector Space Models,

Approaches to Semantics

– Distributional Semantics
  - Compositionality approaches by Marco Baroni’s group:
  - Words are combined with linear matrices dependendent on the P.O.S.:
  - G. Dinu and M. Baroni. How to make words with vectors: Phrase generation in distributional semantics. ACL ’14.

Approaches to Semantics

– Distributional Semantics
  - most recent effort towards solving this problem concern latent factor models because they tend to scale better and to be more robust w.r.t. the heterogeneity of multi-relational data.
  - These models represent entities with latent factors (usually low-dimensional vectors or **embeddings**) and relationships as operators destined to combine those factors.
  - Operators and latent factors are trained to fit the data using reconstruction, clustering or ranking costs.
  - See:
    – Alberto García-Durán, Antoine Bordes, and Nicolas Usunie (2013) Effective Blending of Two and Three-way Interactions for Modeling Multi-relational Data.

Semantic spaces

- Ways of organizing the semantic entities
  - Distributional Semantics
    - Vectors, matrices, tensors
    - Different representations depending on POS
  - Compositional Semantics
    - Atomic units
      - Lexical semantics
    - Complex units
    - Relations between units
      - Ways of composition

Lexical Semantics

- Semantic Dictionaries
- Ontologies
  - Tipology
  - Granularity
  - Scope
  - Genericity
- Examples
  - Domain restricted
    - UMLS, Snomed, BioPortal
  - Generic
    - WordNet, EuroWordnet
  - Other resources

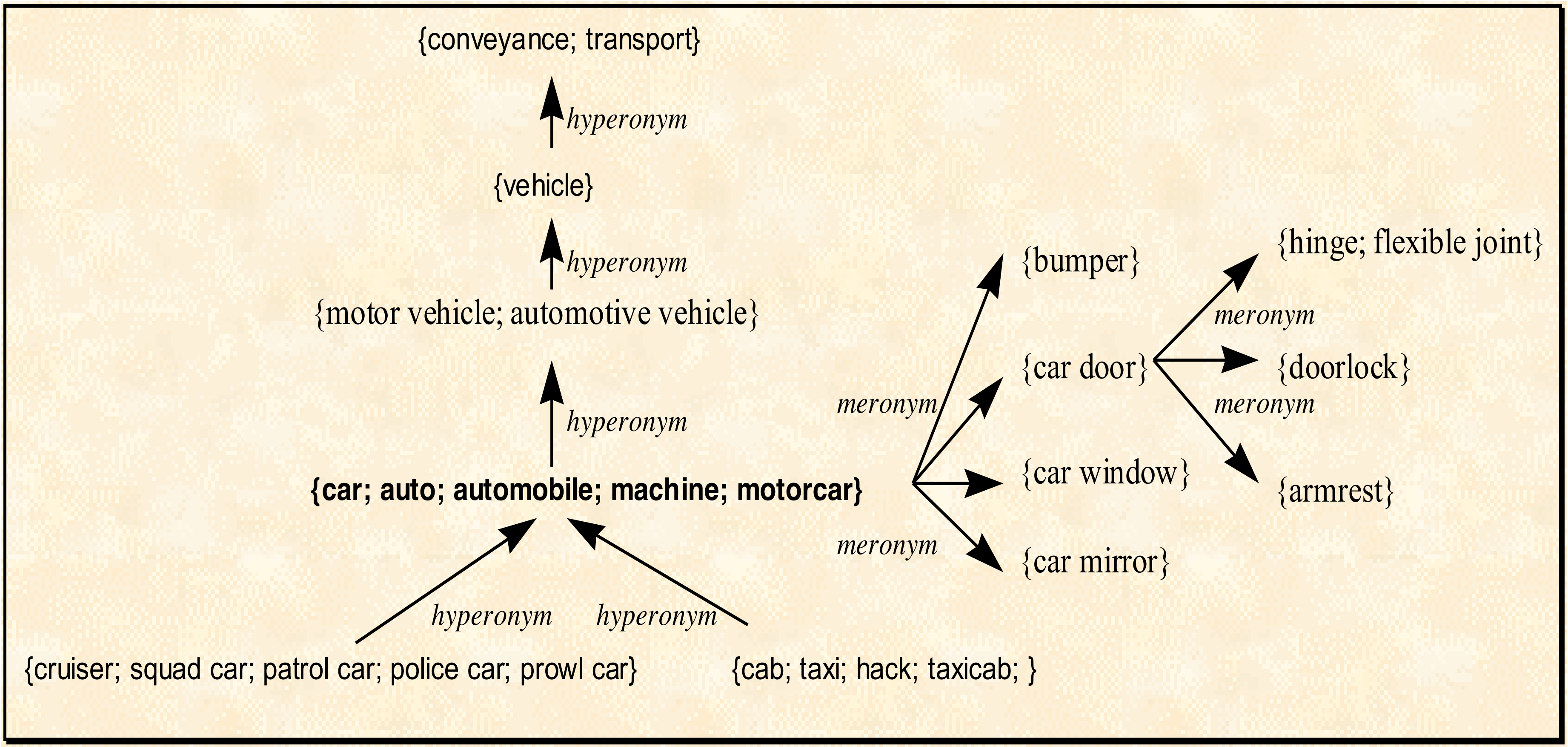
UMLS

- UMLS (Unified Medical Language System)
  - National Library of Medecine, USA Department of Health and Human Services
  - Set of resources
    - Metatesaurus
      - 330.000 concepts, 735.000 terms
    - Semantic Net
      - Basic semantic categories (135 types, 51 relations)
    - Links to vocabularies
      - 30 multilingual sources Lexicón especializa

WordNet

- WordNet
  - Princeton University (Fellbaum,1998)
  - Synsets
  - Nominal, Verbal, Adjectival, Adverbial
  - Semantic relations
    - synonymy
    - antonymy
    - hipernymy-hiponymy
    - meronymy-holonymy
    - entailment
    - cause
    - ...
  - , Extended WordNet

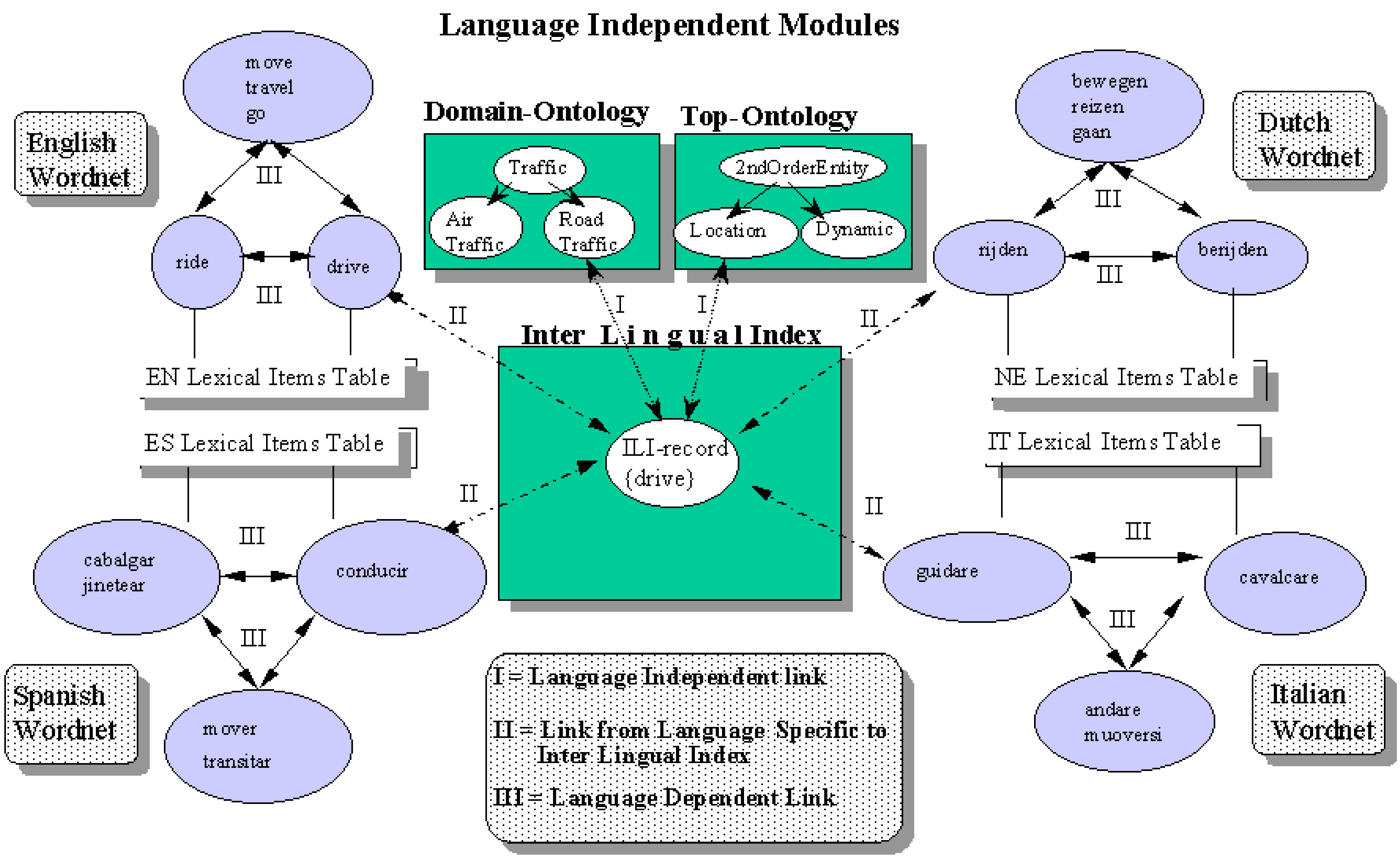
Fragment of WN



Semantic relatedness using WN

- **WordNet::Similarity**
  - **Ted Pedersen**
  - A number of different measures of relatedness have been implemented in this software package. These include a simple edge counting approach and a random method for measuring relatedness. The measures rely heavily on the vast store of knowledge available in the online electronic dictionary -- WordNet.

Architecture of the EuroWordNet Data Structure



Multilingual Central Repository (MCR)

- http://adimen.si.ehu.es/web/MCR/
- The MCR integrates wordnets from five different languages: English, Spanish, Catalan, Basque and Galician. The Inter-Lingual-Index (ILI) allows the connection from words in one language to equivalent translations in any of the other languages thanks to the automatically generated mappings among WordNet versions. The current ILI version corresponds to WordNet 3.0. Furthermore, the MCR is enriched with the semantically tagged glosses.
- The MCR also integrates WordNet Domains, new versions of the Base Concepts and the Top Ontology, and the AdimenSUMO ontology.

Levin classes **(3100 verbs)**

- 47 top level classes, 193 second and third level

- Based on syntactic templares.

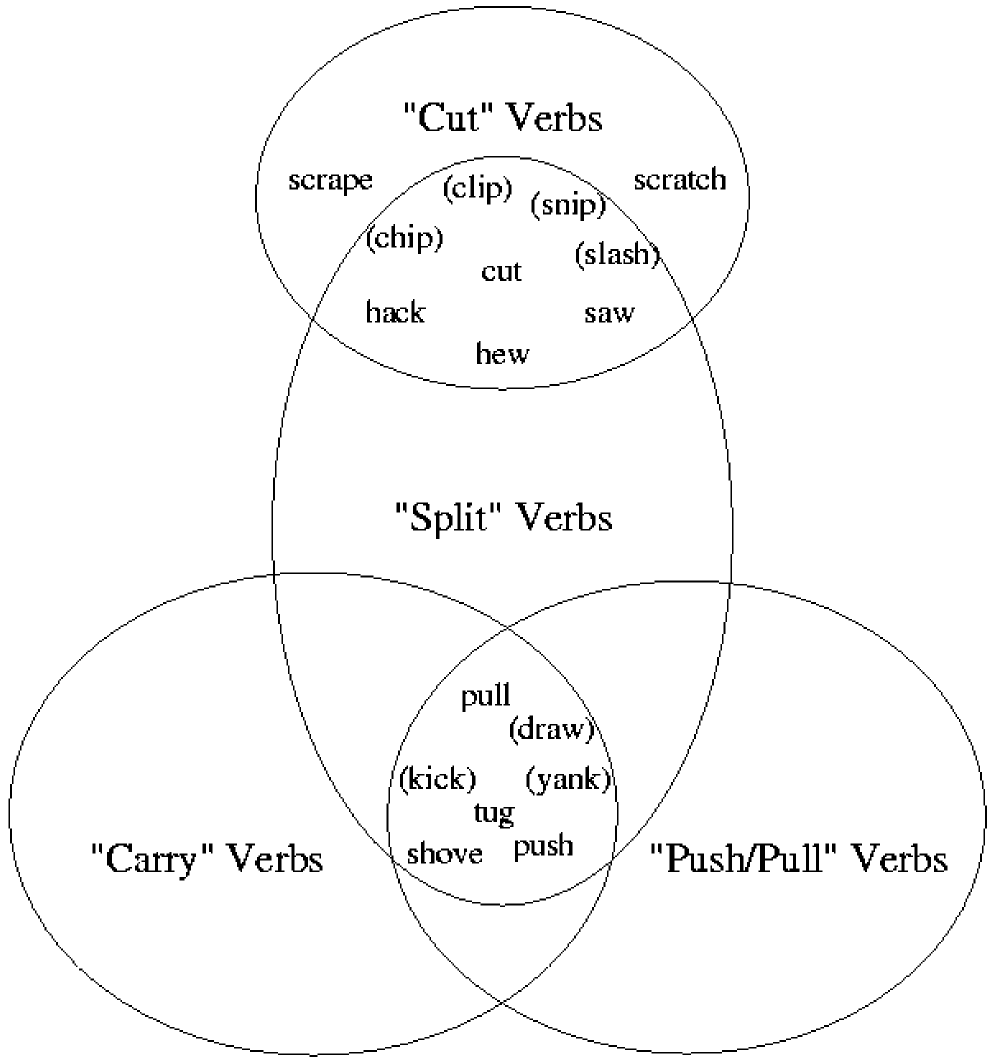
  John broke the jar. / Jars break easily. / The jar broke.
  *John cut the bread. / Bread cuts easily. / *The bread cut.*
  *John hit the wall. / *Walls hit easily. / *The wall hit.*

- They reflect implicitly semantic relations

  contact, directed motion,
  exertion of force, change of state

- Subcategorization templates

Intersective Levin classes



VerbNet

- From Intersective Levin Classes
  - More syntactically and semantically coherent
  - sets of syntactic patterns
  - explicit semantic components
  - relations between senses

- VERBNET
  - verbs.colorado.edu/verb-index/index.php
  - Martha Palmer et al.

VerbNet

- Class entries:
  - Capture generalizations about verb behavior
  - Organized hierarchically
  - Members have common semantic elements, **semantic roles (28**) and syntactic frames
- Verb entries:
  - Refer to a set of classes (different senses)
  - each class member linked to WN synset(s) and FrameNet frames
  - Currently 6,300 verbs

VerbNet

- Organizes verbs into classes that have common syntax/semantics linking behavior
- Classes include…
  - A list of member verbs (w/ WordNet senses)
  - A set of thematic roles (w/ selectional restr.s)
  - A set of frames, which define both syntax & semantics using thematic roles.
- Classes are organized hierarchically

VerbNet Thematic Roles

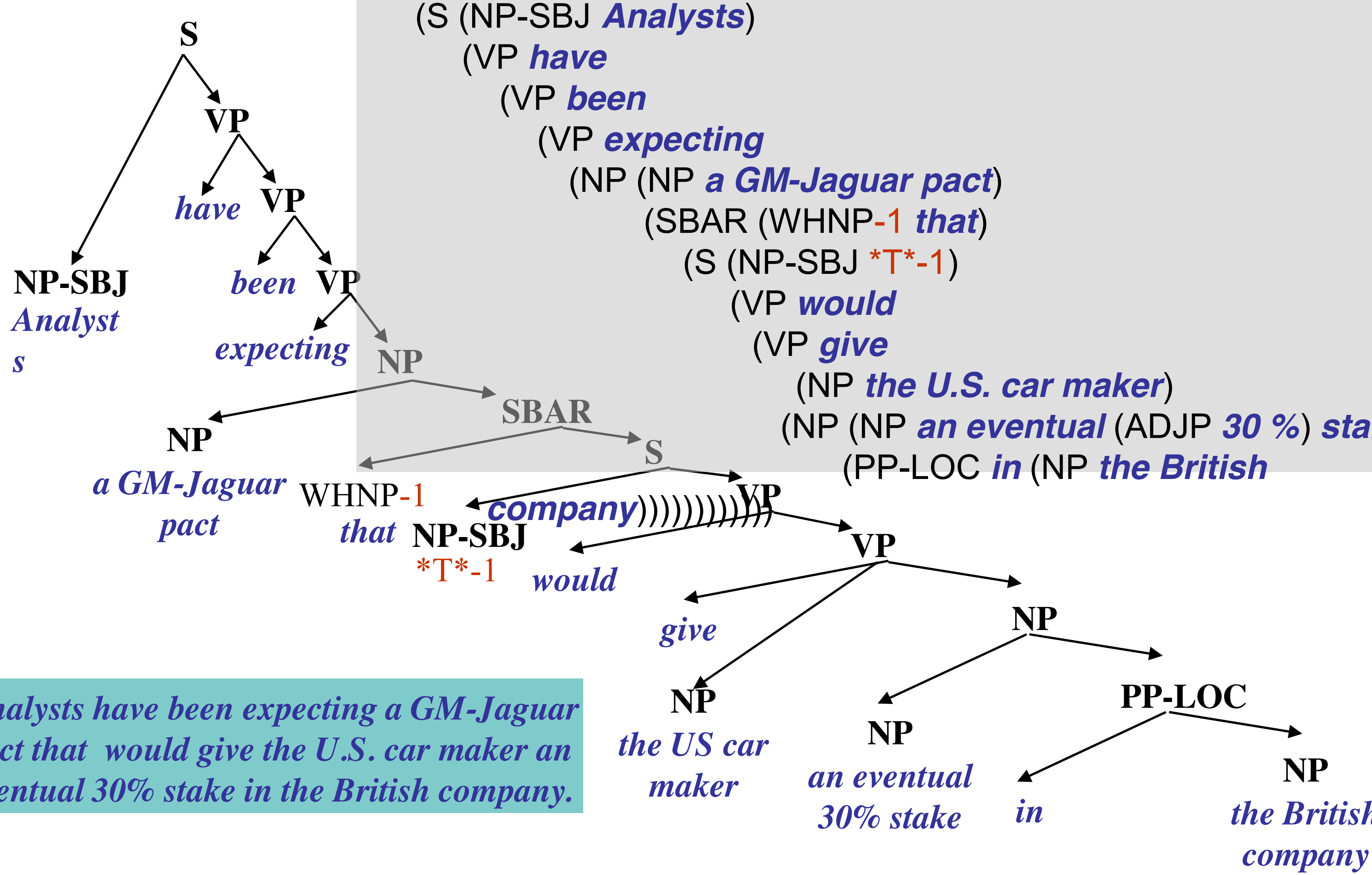
- Actor
- Actor1
- Actor2
- Agent
- Asset
- Attribute
- Beneficiary
- Cause
- Destination
- Experiencer
- Extent
- Instrument
- Location
- Material
- Patient
- Patient1
- Patient2
- Predicate
- Product
- Proposition
- Recipient
- Source
- Stimulus
- Theme
- Theme1
- Theme2
- Time
- Topic
- Value

Penn Treebank

- 1.3 Mw, 40,000 sentences
- Wall Street Journal and other sources
- POS tagged
- Syntactically Parsed

A TreeBanked Sentence

(S (NP-SBJ ***Analysts***)
(VP ***have***
(VP ***been***
(VP ***expecting***
(NP (NP ***a GM-Jaguar pact***)
(SBAR (WHNP-1 ***that***)
(S (NP-SBJ *T*-1)
(VP ***would***
(VP ***give***
(NP ***the U.S. car maker***)
(NP (NP ***an eventual*** (ADJP ***30 %***) ***stake***)
(PP-LOC ***in*** (NP ***the British company***))))))))))))



Analysts have been expecting a GM-Jaguar pact that would give the U.S. car maker an eventual 30% stake in the British company.

Proposition Bank (Propbank)

Generalization from sentences to propositions

Powell met Zhu Rongji

Powell and Zhu Rongji met

Powell met with Zhu Rongji

Powell and Zhu Rongji had a meeting

. . .

battle

wrestle

join

debate

consult

Proposition: **meet**(Powell, Zhu Rongji)

meet(Somebody1, Somebody2)

When Powell met Zhu Rongji on Thursday they discussed the return of the spy plane.

meet(Powell, Zhu) **discuss([Powell, Zhu], return(X, plane))**

PropBank

- 1M words of WSJ annotated with predicate-argument structures for verbs.
  - The location & type of each verb's arguments
- Argument types are defined on a per-verb basis.
  - Consistent across uses of a single verb (sense)
- But the same tags are used (Arg0, Arg1, Arg2, …)
  - Arg0 ≈ proto-typical agent (Dowty)
  - Arg1 ≈ proto-typical patient

PropBank

- *Example: cover (smear, put over)*
- Arguments:
  - Arg0 = causer of covering
  - Arg1 = thing covered
  - Arg2 = covered with
- Example:

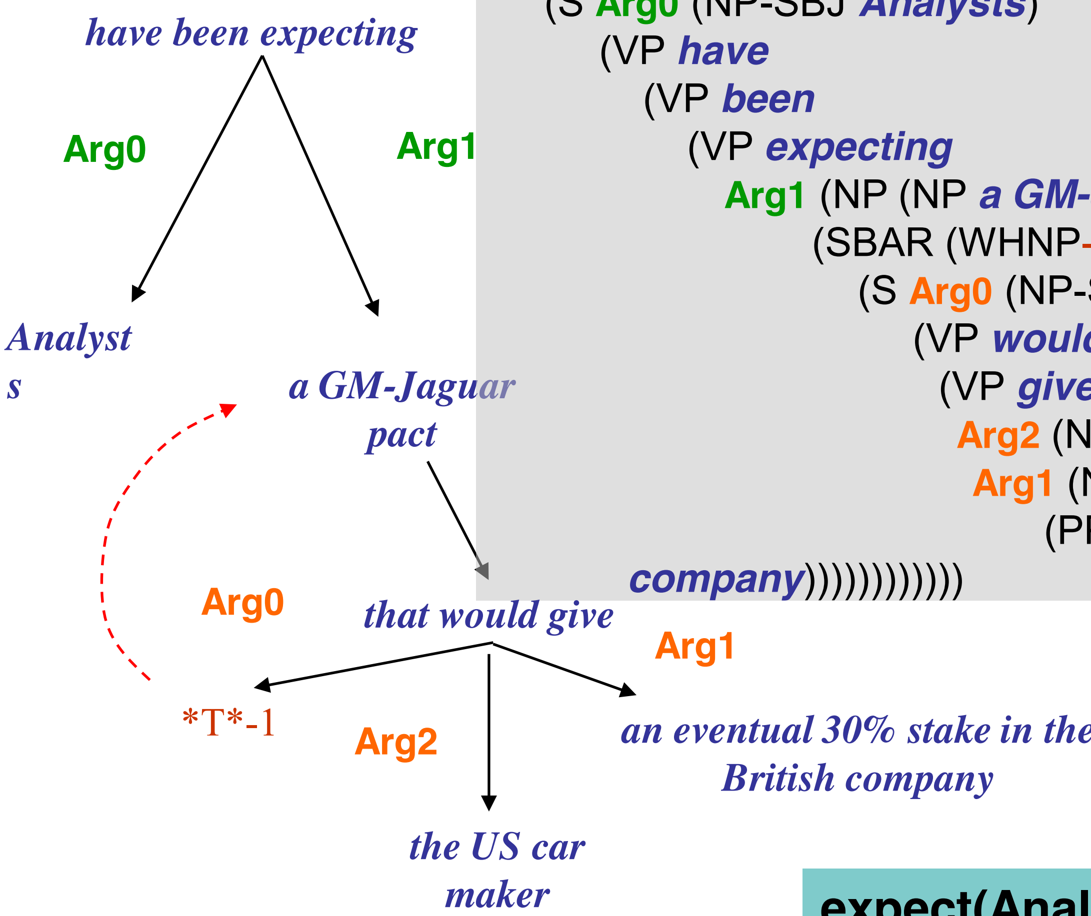
  John ***covered*** the bread with peanut butter.

PropBank

- Trends in Argument Numbering
- Arg0 = proto-typical agent (*Dowty*)
  Agent (85%), Experiencer (7%), Theme (2%), …
- Arg1 = proto-typical patient
  Theme (47%),Topic (23%), Patient (11%), …
- Arg2 = Recipient (22%), Extent (15%), Predicate (14%), …
- Arg3 = Asset (33%), Theme2 (14%), Recipient (13%), …
- Arg4 = Location (89%), Beneficiary (5%), …
- Arg5 = Location (94%), Destination (6%)

(Percentages indicate how often argument instances were mapped to VerbNet roles in the PropBank corpus)

The same sentence, PropBanked




```
(S Arg0 (NP-SBJ Analysts)
    (VP have
       (VP been
          (VP expecting
             Arg1 (NP (NP a GM-Jaguar pact)
                   (SBAR (WHNP-1 that)
                      (S Arg0 (NP-SBJ *T*-1)
                         (VP would
                           (VP give
                              Arg2 (NP the U.S. car maker)
                              Arg1 (NP (NP an eventual (ADJP 30 %) stake)
                                   (PP-LOC in (NP the British
      company))))))))))))
```

expect(Analysts, GM-J pact)
give(GM-J pact, US car maker, 30% stake)

FrameNet

- http://framenet.ICSI.berkeley.edu/framenet
- [Baker, Sato, 2003], [C.F. Baker, C.J. Fillmore, and J.B. Lowe. 1998]
- Semantic frame
  - type of event or state and the participants and “props” associated with it:
- frame element (FE)
- Frames range from highly abstract to quite specific. An example of an abstract frame would be the Replacement frame, with Fes such as OLD and NEW:
  - Pat replaced [$_{Old}$ the curtains] [$_{New}$ with wooden blinds]
- One sense of the verb replace is associated with the Replacement frame, thus constituting one lexical unit
- Lexical Unit (LU), the basic unit of the FrameNet lexicon.
- An example of a more specific frame is Apply_heat, with FEs such as COOK, FOOD, MEDIUM, and DURATION
  - Boil [$_{Food}$ the rice] [$_{Duration}$ for 3 minutes] [$_{Medium}$ in water], then drain

NomBank

- http://nlp.cs.nyu.edu/meyers/NomBank.html
- NomBank is an annotation project at New York University that is related to the PropBank project at the University of Colorado
  - A. Meyers, et al, 2004
- NomBank will provide argument structure for instances of about 5,000 common nouns in the Penn Treebank II corpus.
  - PropBank:
    - REL = gave, ARG0 = they, ARG1 = a standing ovation, ARG2 = the chefs
  - NomBank:
    - REL = ovation, ARG0 = they, ARG1 = the chefs, SUPPORT = gave
- NomBank.1.0
  - covering all the "markable" nouns in the PTB-II WSJ corpus.
  - 114,576 propositions derived from looking at a total of 202,965 noun instances and choosing only those nouns whose arguments occur in the text.

Putting all together

- Loper, Yi, Palmer, 2006
  - PropBank
    - How does a verb relate to its arguments? Includes annotated text.
  - VerbNet
    - How do verbs w/ shared semantic & syntactic features(and their arguments) relate?
  - FrameNet
    - How do verbs that describe a common scenario relate?
  - WordNet
    - What verbs are synonymous?
  - Cyc
    - How do verbs relate to a knowledge based ontology?
- => SemLink

Putting all together

- In PropBank, Arg2-Arg5 are overloaded.
  - But in VerbNet, the same thematic roles acrossverbs.
- PropBank training data is too domain specific.
- =>
  - Use VerbNet as a bridge to merge PropBank w/FrameNet
  - Expand the size and variety of the training data

Putting all together

- Abstract Meaning Representations – AMR
- Knight, et. al., LAW-2013
- Example:
  - *He was not aware of research on smokers of the Kent cigarettes.*

```
(r / realize-01
     :polarity -
     :ARG0 (h / he)
     :ARG1 (r3 / research-01
                :ARG1 (p4 /person
                          :ARG0-of (s / smoke-02
                                      :ARG1 (c2 / cigarette
                                                :name (k / name
                                                         op1: Kent)))))
```

DIRT

- **DIRT Paraphrase Collection**
- **DIRT** (Discovery of Inference Rules from Text) is both an algorithm and a resulting knowledge collection
  - Dekang Lin and Patrick Pantel (2001)
  - A path, extracted from a dependency parse tree, is an expression that represents a binary relationship between two nouns. If two paths tend to link the same sets of words, DIRT hypothesizes that the meanings of the corresponding patterns are similar.
- The DIRT knowledge collection
  - 7 million paths from the parse trees (231,000 unique) from which scored paraphrases were generated. Here are the top paraphrases "X solves Y" generated by DIRT:
    - *Y is solved by X, X resolves Y, X finds a solution to Y, X tries to solve Y, X deals with Y, Y is resolved by X, X addresses Y, …*

DART

- **DART Database**
  - http://www.cs.utexas.edu/users/pclark/dart/
  - P. Clark, P. Harrison, 2009
  - The DART (Discovery and Aggregation of Relations in Text) database contains approximately 23 million distinct "world knowledge" propositions (110 million with duplicates), extracted from text by abstracting parse trees.
  - 12 kinds of proposition, contained in 12 different text files

DART

- **Frequency Tuple Proposition Verbalization**

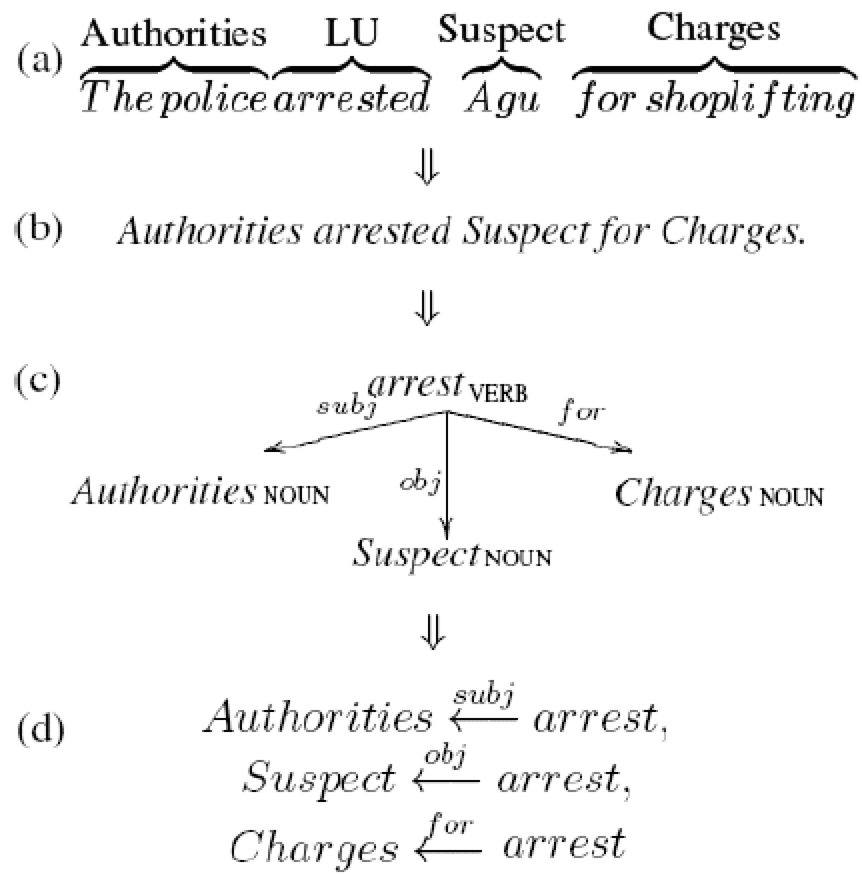
Frequency	Tuple Proposition	Verbalization
144	(an "small" "hotel")	"Hotels can be small.“
121	(anpn "subject" "agreement" "to" "approval")	"Agreements can be subject to approvals."
17	(nn "drug" "distributor")	"There can be drug distributors."
153	(nv "bus" "carry")	"Buses can carry [something/someone]."
26	(npn "sentence" "for" "offence")	"Sentences can be for offences."
119	(nvn "critic" "claim" "thing")	"Critics can claim things."
192	(nvpn "person" "go" "into" "room")	"People can go into rooms."
11	(nvnpn "democrat" "win" "seat" "in" "election")	"Democrats can win seats in elections."
1572	(qn "year" "contract")	"Contracts can be measured in years."
8	(vn "find" "spider")	"Spiders can be found."
14	(vpn "refer" "to" "business")	"Referring can be to businesses."
103	(vnpn "educate" "person" "at" "college")	"People can be educated at colleges."

REVERB

- Predicative entailment rules contains three resources in two formats – shallow and syntactic. Resources are learned over the REVERB data set and using the local and algorithms described in Chapter 5 of Jonathan Berant's thesis
- REVERB data set contains 103,315 distinct predicates, which appear with a large number of distinct arguments and pairs of arguments.
- Every pair of predicates is represented by a feature vector
- Ex. X defeat Y => Y lose to X

FRED

- **FRED – FrameNet-derived entailment rule-base**
  - Roni Ben Aharon, Idan Szpektor and Ido Dagan. ACL 2010.
  - http://www.cs.biu.ac.il/~nlp/downloads

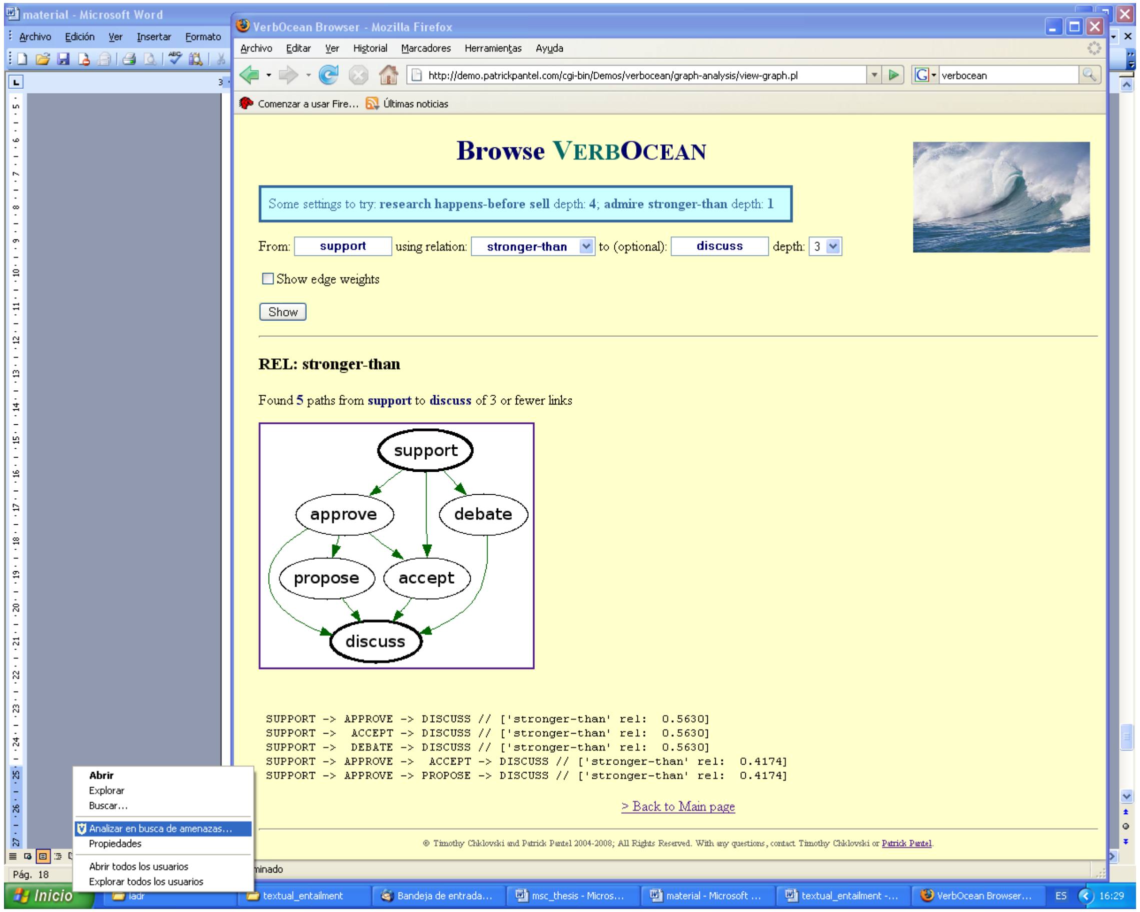


Ancora

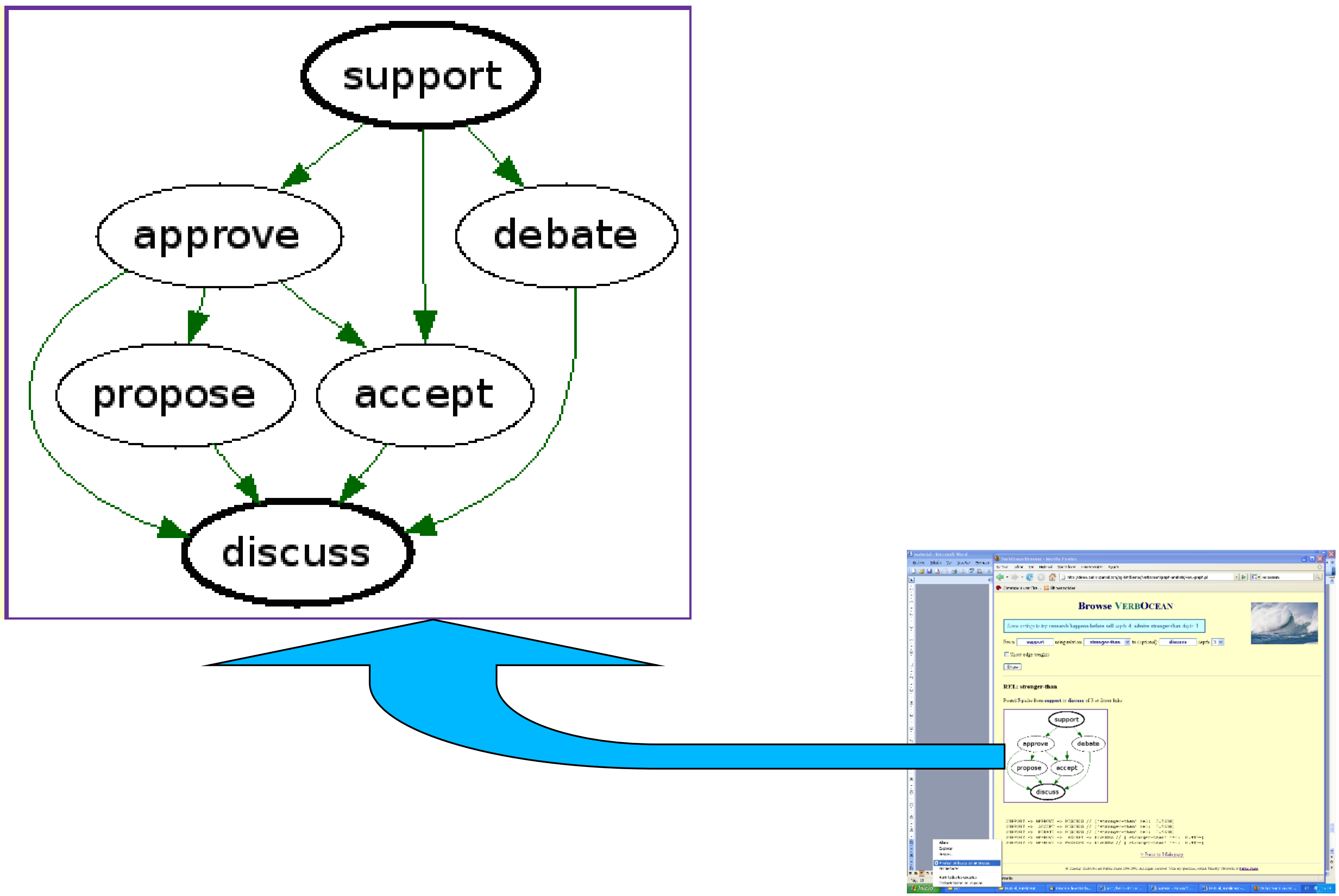
- Treebank of Spanish and Catalan
- University of Barcelona
- 0.5 Mw
- Constituent & dependency parsed
- Coreference tagged
- WN synsets tagged
- Role labels explicit & implicit
- Ancora-verb
- Ancora-nom

VERBOCEAN

- Timothy Chklovski and Patrick Pantel (2004)
- http://semantics.isi.edu/ocean/.

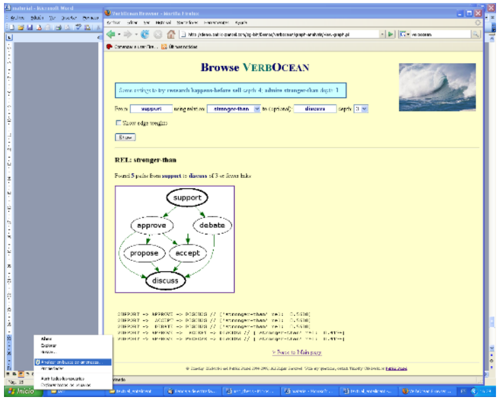


VERBOCEAN



VERBOCEAN

```
SUPPORT -> APPROVE -> DISCUSS // ['stronger-than' rel:  0.5630]
SUPPORT ->  ACCEPT -> DISCUSS // ['stronger-than' rel:  0.5630]
SUPPORT ->  DEBATE -> DISCUSS // ['stronger-than' rel:  0.5630]
SUPPORT -> APPROVE ->  ACCEPT -> DISCUSS // ['stronger-than' rel:  0.4174]
SUPPORT -> APPROVE -> PROPOSE -> DISCUSS // ['stronger-than' rel:  0.4174]
```



Wikipedia

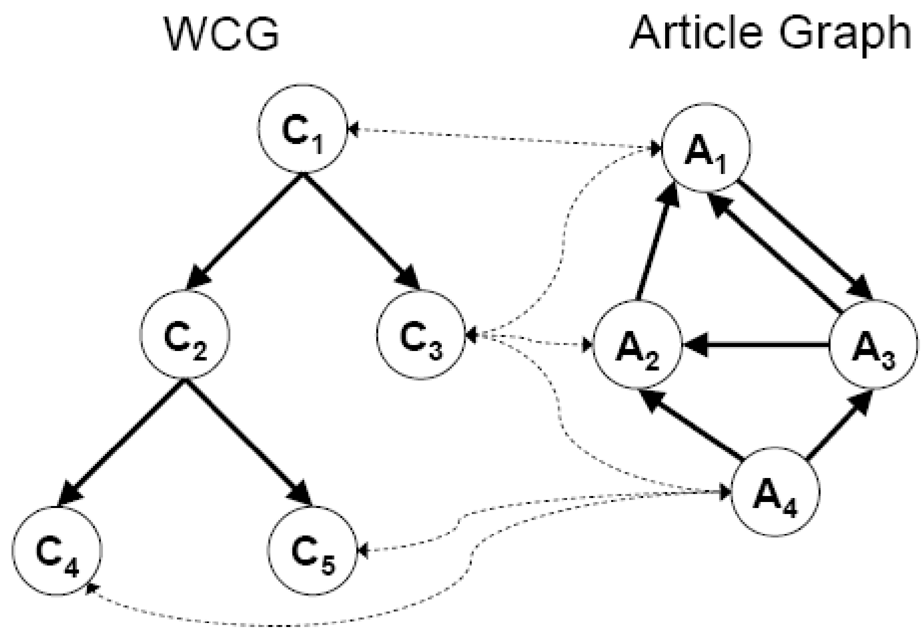
- More than 300 languages
- More than 32M pages
  - English > 4M pages
  - 8 languages with > 1M pages
- Survey of applications in Medelyan et al, 2009

Organization of Wikipedia

- Types of links
  - *Article links*
    - links from one article to another of the same language;
  - *Category links*
    - links from an article to special “Category” pages;
  - *Interlingual links*
    - links from an article to a presumably equivalent, article in another language;
- Types of special pages
  - *Redirect pages*
    - short pages which often provide equivalent names for an entity
  - *Disambiguation pages*
    - a page with little content that links to multiple similarly named articles.
- Infoboxes, templates, list pages, wikipedia commons, ...

Organization of Wikipedia

- Torsten Zesch and Iryna Gurevych, 2007
  - Wikipedia Article Graph, WAG
  - Wikipedia Category Graph, WCG



Accessing Wikipedia

- Iryna Gurevych's JWPL software
  - https://www.ukp.tu-darmstadt.de/software/jwpl/
  - Torsten Zesch and Christof Müller and Iryna Gurevych, 2008
  - JWPL (Java Wikipedia Library) is a open-source, Java-based application programming interface that allows to access all information contained in Wikipedia. The high-performance Wikipedia API provides structured access to information nuggets like redirects, categories, articles and link structure.
- Using python wikitools
  - https://code.google.com/p/python-wikitools/
  - Python package to interact with the MediaWiki API. The package contains general tools for working with wikis, pages, and users on the wiki and retrieving data from the MediaWiki API.

Related and Derived Resources

- DBpedia
  - U. Leipzig, Freie U. Berlin
  - Auer at al, 2007
  - Interlinking DBpedia with other datasets:
    - Geonames, WordNet, OpenCyc, FreeBase, ...
  - Sparql dbpedia endpoint
    - http://dbpedia.org/sparql
- Wikipedia XML corpus
- Yago, later Yago 2
  - Suchanek, 2008
  - Suchanek et al, 2007
- Semantic Wikipedia
  - Max Völkel at al, 2008
- Yahoo's Correlator
  - Yahoo's Barcelona Media Research Center
- Linking WP to ResearchCyc ontology
  - Medelyan, Legg, 2008

Related and Derived Resources

Processing tools	
JWPL Java Wikipedia Library	API for structural access of Wikipedia parts such as redirects, categories, articles and link structure. [Zesch et al. 2008] *http://www.ukp.tu-darmstadt.de/software/jwpl/*
WikiRelate!	API for computing semantic relatedness using Wikipedia [Strube and Ponzetto 2006; Ponzetto and Strube 2006] *http://www.eml-research.de/ english/research/ nlp/download/ wikipediasimilarity.php*
Wikipedia Miner	API that provides a simplified access to Wikipedia and models its structure semantically [Milne et al. 2008] *http://sourceforge.net/ projects/wikipedia-miner/*
WikiPrep	A Perl tool for preprocessing Wikipedia XML dumps [Gabrilovich and Markovitch 2007] *http://www.cs.technion.ac.il/ ~gabr/resources/ code/wikiprep/*
W.H.A.T. Wikipedia Hybrid Analysis Tool	An analytic tool for Wikipedia with two main functionalities: an article network and extensive statistics. It contains a visualization of the article networks and a powerful interface to analyze the behavior of authors. *http://sourceforge.net/ projects/ w-h-a-t/*

Related and Derived Resources

Wikipedia mining demos	
DBpedia Online Access	Online access of DBpedia data (103M facts extracted from Wikipedia) via a SPARQL query endpoint and as Linked Data. [Auer et al. 2007] *http://wiki.dbpedia.org/ OnlineAccess*
YAGO	Demo of the Yet Another Ontology YAGO, containing 1.7M entities and 14M facts [Suchanek et al. 2007] *http://www.mpii.mpg.de/ ~suchanek/yago*
QuALiM	A Question Answering system. Given a question in a natural language returns relevant passages from Wikipedia. [Kaisser 2008] *http://demos.inf.ed.ac.uk:8080/ qualim/*
Koru	A demo of a search interface that maps topics involved in both queries and documents to Wikipedia articles. Supports automatic and interactive query expansion. [Milne et al. 2007] *http://www.nzdl.org/koru*
Wikipedia Thesaurus	A large scale association thesaurus containing 78 million associations [Nakayama et al. 2007 and 2008] *http://wikipedia-lab.org:8080/ WikipediaThesaurusV2/*

DBPedia

Dataset	Description	Triples
Page links	Internal links between DBpedia instances derived from the internal pagelinks between Wikipedia articles	62 M
Infoboxes	Data attributes for concepts that have been extracted from Wikipedia infoboxes	15.5 M
Articles	Descriptions of all 1.95 million concepts within the English Wikipedia. Includes titles, short abstracts, thumbnails and links to the corresponding articles	7.6 M
Languages	Additional titles, short abstracts and Wikipedia article links in 13 other languages.	5.7 M
Article categories	Links from concepts to categories using SKOS	5.2 M
Extended abstracts	Additional, extended English abstracts	2.1 M
Language abstracts	Extended abstracts in 13 languages	1.9 M
Type information	Inferred from category structure and redirects by the YAGO ("yet another great ontology") project [Suchanek et al. 2007]	1.9 M
External links	Links to external web pages about a concept	1.6 M
Categories	Information which concept is a category and how categories are related	1 M
Persons	Information about 80,000 persons (date and place of birth etc.) represented using the FOAF vocabulary	0.5 M
External links	Links between DBpedia and Geonames, US Census, Musicbrainz, Project Gutenberg, the DBLP bibliography and the RDF Book Mashup	180 K

Table 6. Content of DBPedia [Auer et al. 2007].

Accessing dbpedia through virtuoso endpoint

- Sparql query:
  - select distinct ?Concept where {[] a ?Concept} LIMIT 10

- Concept
  - http://www.w3.org/2004/02/skos/core#Concept http://xmlns.com/foaf/0.1/Person http://schema.org/Person http://wikidata.dbpedia.org/resource/Q215627 http://www.w3.org/2002/07/owl#Thing http://wikidata.dbpedia.org/resource/Q5 http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#Agent http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#NaturalPerson http://dbpedia.org/ontology/Agent http://dbpedia.org/ontology/Athlete

Semantic Wikipedia

Using Wikipedia in NLP tasks and Applications

- Measures of semantic relatedness using Wikipedia
  - Strube and Ponzetto, 2006
  - Gabrilovich and Markovitch, 2007
  - Torsten Zesch and Iryna Gurevych, 2007
  - Milne and Witten, 2008
- Named Entities Recognition & Classification
  - Bunescu and Pasca, 2006
  - Cucerzan, 2007
  - Buscaldi, Rosso, 2007
- Word Sense Disambiguation
  - Mihalcea, 2008
  - Mihalcea, Csomai, 2007
- Extending existing thesauri

Using Wikipedia in NLP tasks and Applications

- Semantic tagging & Topic indexing
  - Milne and Witten, 2008
  - Medelyan et al, 2008
  - Mihalcea and Csomai (2007)
  - Atserias et al, 2008
  - Milka et al, 2008
  - Wu, Weld, 2007, 2008
- Improving IR access to documents
  - Milne et al, 2007
- Text categorization
  - Gabrilovich and Markovitch, 2007
  - Gabrilovich, 2006

Using Wikipedia in NLP tasks and Applications

- Multilingual applications
  - Ferrández et al, 2007
  - Alexander E. Richman and Patrick Schone, 2008
  - Erdmann et al, 2008
- Coreference Resolution
  - Ponzetto and Strube, 2008
- Mining Domain-Specific Thesauri
  - Milne et al, 2006
- Q&A
  - QuALiM, Kaisser, 2008
- Textual Entailment
  - Eyal Shnarch, 2008

Measures of semantic relatedness using Wikipedia

Method	M&C	R&G	WS-353
WordNet [Strube and Ponzetto, 2006]	0.82	0.86	full: 0.36 test: 0.38
WikiRelate! [Ponzetto and Strube, 2007]	0.49	0.55	full: 0.49 test: 0.62
ESA [Gabrilovich and Markovitch, 2007]	0.73	0.82	0.75
WLVM [Milne, 2007]	n/a	n/a	man: 0.72 auto: 0.45
WLM [Milne and Witten, 2008]	0.70	0.64	0.69

Table 2. Overview of semantic relatedness methods.

Measures of semantic relatedness using Wikipedia

- Measures of semantic relatedness using Wikipedia
  - Strube and Ponzetto, 2006
  - Gabrilovich and Markovitch, 2007
  - Torsten Zesch and Iryna Gurevych, 2007
  - Milne and Witten, 2008

Freebase

- Freebase
  - https://www.freebase.com/
  - Freebase is an open database of the world's information. It is built by the community and for the community—free for anyone to query, contribute to, built applications on top of, or integrate into their websites
  - Freebase has an RDF service that exposes URIs and generates RDF descriptions for all Freebase topics.
  - **2,751,750,754** Facts
  - **47,433,069** Topics

Freebase

- Freebase topics & facts
  - Music 31M 213M
  - Books 6M 15M
  - Media 5M 17M
  - People 3M 20M
  - …

Lexical Semantics Tasks

- WSD
- NEC
- Semantic tagging
  - Wikification
- Terminology detection
- MWE detection & classification
- Entity Linking (grounding)
- GeoDisambiguation, GeoLocalisation, GeoReferencing, Placing

Word Sense Disambiguation (WSD)

- Sense
  - distinction of meaning of a word (word type) occurring in different mentions (word tokens)
- Given a mention which is its correct sense
- Sense tagsets:
  - WN, WP, Clusters of words
- Surveys:
  - Agirre, E., Edmonds, P. (2006) : Word sense disambiguation: Algorithms and applications AAAI Workshop (2006)
  - Navigli, R. (2009): Word sense disambiguation: A survey. In: ACM Comput. Surveys, Volume 41, (2009)
  - Gerber, A., Gao, L., Hunte, J. (2011): A scoping study of (who, what, when, where) semantic tagging services. In: Research Report, eResearch Lab, The University of Queensland, (2011).
  - Moro, A., Roganato, A., Navigli, R. . (2014) : Entity linking meets word sense disambiguation: A unied approach. In: Transactions of ACL (2014)

- Semantic parsing includes performing word sense disambiguation

State? River?

Which rivers run through the states bordering Mississippi?

Semantic Parsing

answer(traverse(next_to(stateid('mississippi'))))

WSD

- Frequent Restrictions
  - Yarowsky (1995)
    - One sense per discourse
    - One sense per collocation

WSD

- A baseline solutiion (Naive Bayes)
  - Let **w** the word to disambiguate
  - Let c_k the possible senses
  - Let $\vec{x}$ the context vector (e.g. fixed window of 100 words).

$$c' = \arg\max_{c_k} P(c_k \mid \vec{x})$$

  - Applying Bayes we have:

WSD

$$c' = \arg\max_{c_k} P(c_k \mid \vec{x}) =$$

$$\arg\max_{c_k} \frac{P(\vec{x} \mid c_k)}{P(\vec{x})} \cdot P(c_k) =$$

$$\arg\max_{c_k} P(\vec{x} \mid c_k) \cdot P(c_k) =$$

$$\arg\max_{c_k} \left[\log P(\vec{x} \mid c_k) + \log P(c_k)\right] =$$

$$\arg\max_{c_k} \left[\sum_{v_j \ in \ \vec{x}} \log P(v_j \mid c_k) + \log P(c_k)\right]$$

Semantic tagging

- Milne and Witten, 2008
  - there are 26 possible senses. Only one sense is a positive example, and the remaining 25 are negative. In all, the 500 training articles provide about 1.8 million examples.

Depth-first search

From Wikipedia, the free encyclopedia

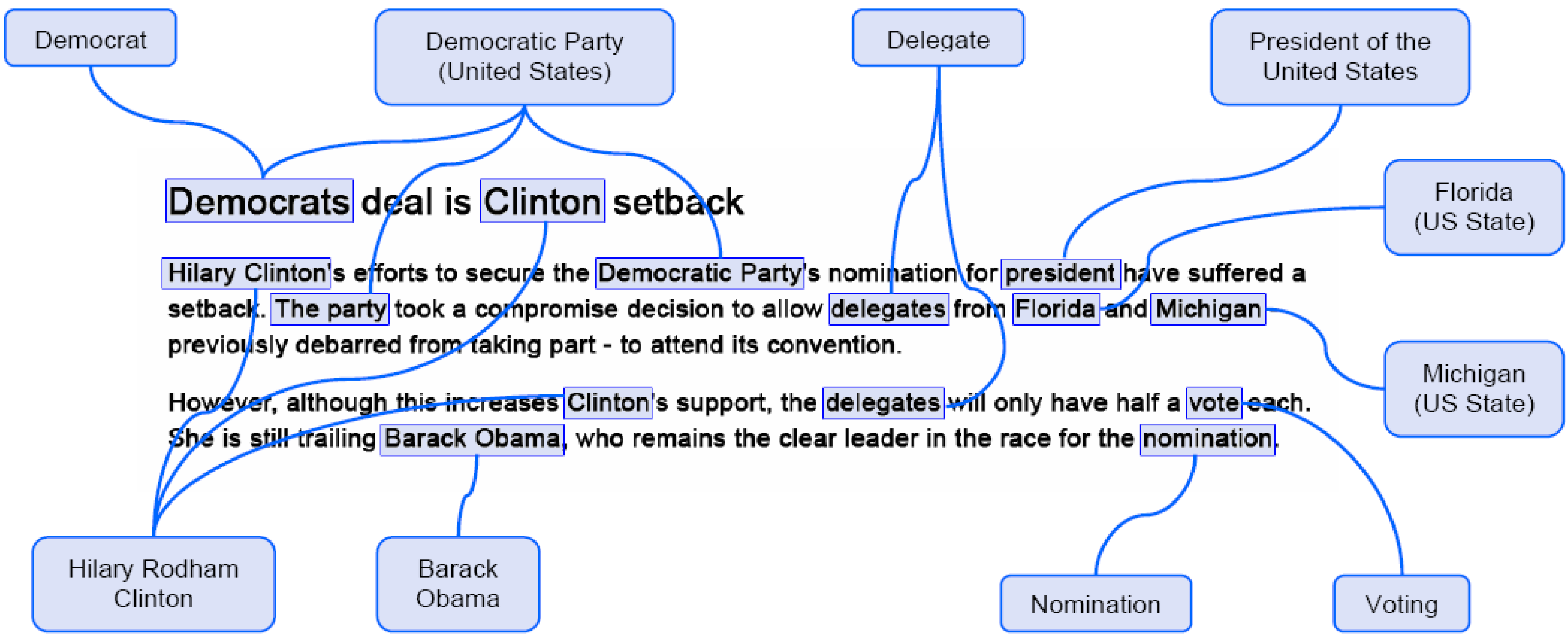
Depth-first search (**DFS**) is an algorithm for traversing or searching a tree, tree structure, or graph. One starts at the root (selecting some node as the root in the graph case) and explores as far as possible along each branch before backtracking.

Formally, DFS is an uninformed search that progresses by expanding the first child node of the search tree that appears and thus going deeper and deeper until a goal node is found, or until it hits a node that has no children. Then the search backtracks, returning to the most recent node it hadn't finished exploring. In a non-recursive implementation, all freshly expanded nodes are added to a LIFO stack for exploration.

sense	commonness	relatedness
Tree	92.82%	15.97%
Tree (graph theory)	2.94%	59.91%
Tree (data structure)	**2.57%**	**63.26%**
Tree (set theory)	0.15%	34.04%
Phylogenetic tree	0.07%	20.33%
Christmas tree	0.07%	0.0%
Binary tree	0.04%	62.43%
Family tree	0.04%	16.31%
...		

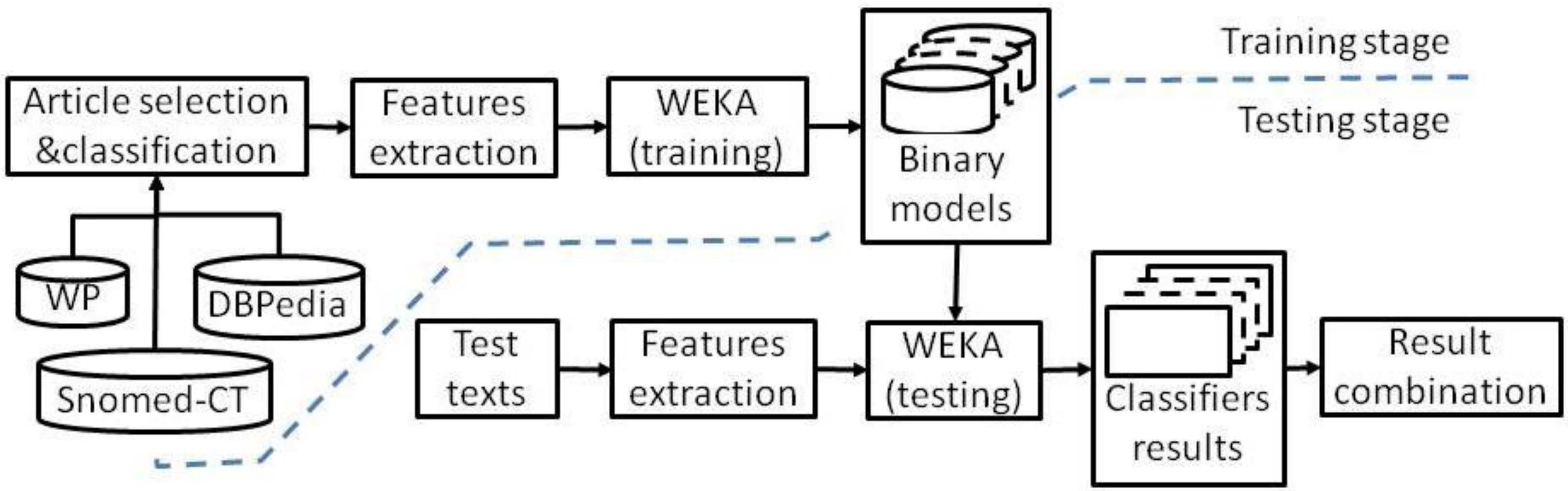
Semantic tagging

- Milne and Witten, 2008



Semantic tagging

- An example:
  - ST in the medical domain
    - Vivaldi, Rodríguez, 2015



Wikifiers

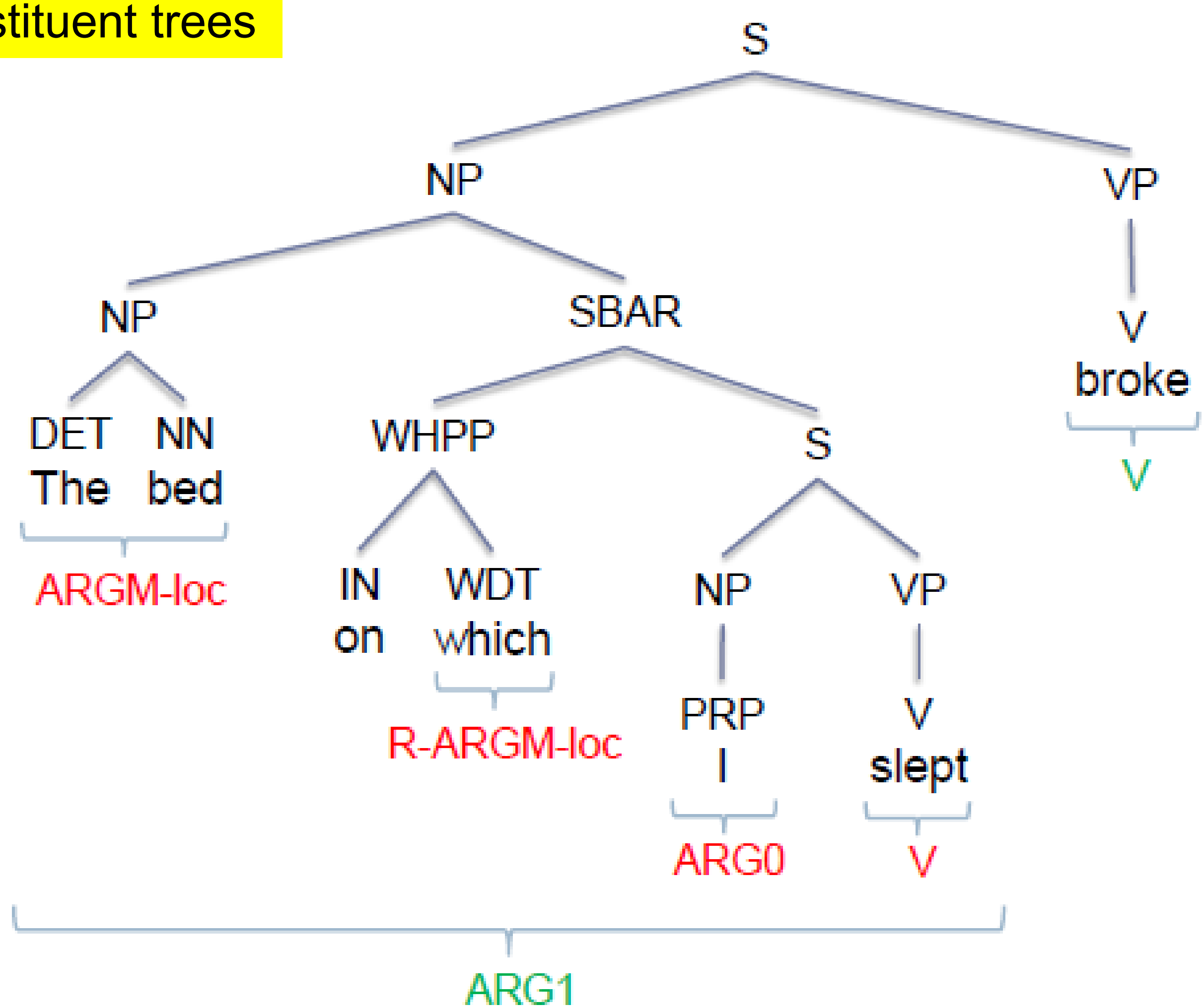
- Seminal works:
  - Mihalcea & Csomai: Wikify!,
  - Cucerzan
  - Milne & Witten
- Recent systems:
  - CSAW
  - Illinois Wikifier
  - TAGME
  - DBPedia Spotlight,
  - AIDA
  - RPI Wikifier
- Most of these systems proceed into two steps:
  - candidate detection
  - classification or ranking

Semantic Role Labeling

- SRL:
  - Semantic Role Labeling Tutorial, NAACL, June 9, 2013
    - Martha Palmer, Shumin Wu, Ivan Titov
  - Capturing Semantic Roles
  - Predicates + Arguments
    - Predicates realized as verbs or nominalizations
    - Explicit or Implicit Arguments
  - Role definitions have to be determined mention by mention, and with respect to the other roles

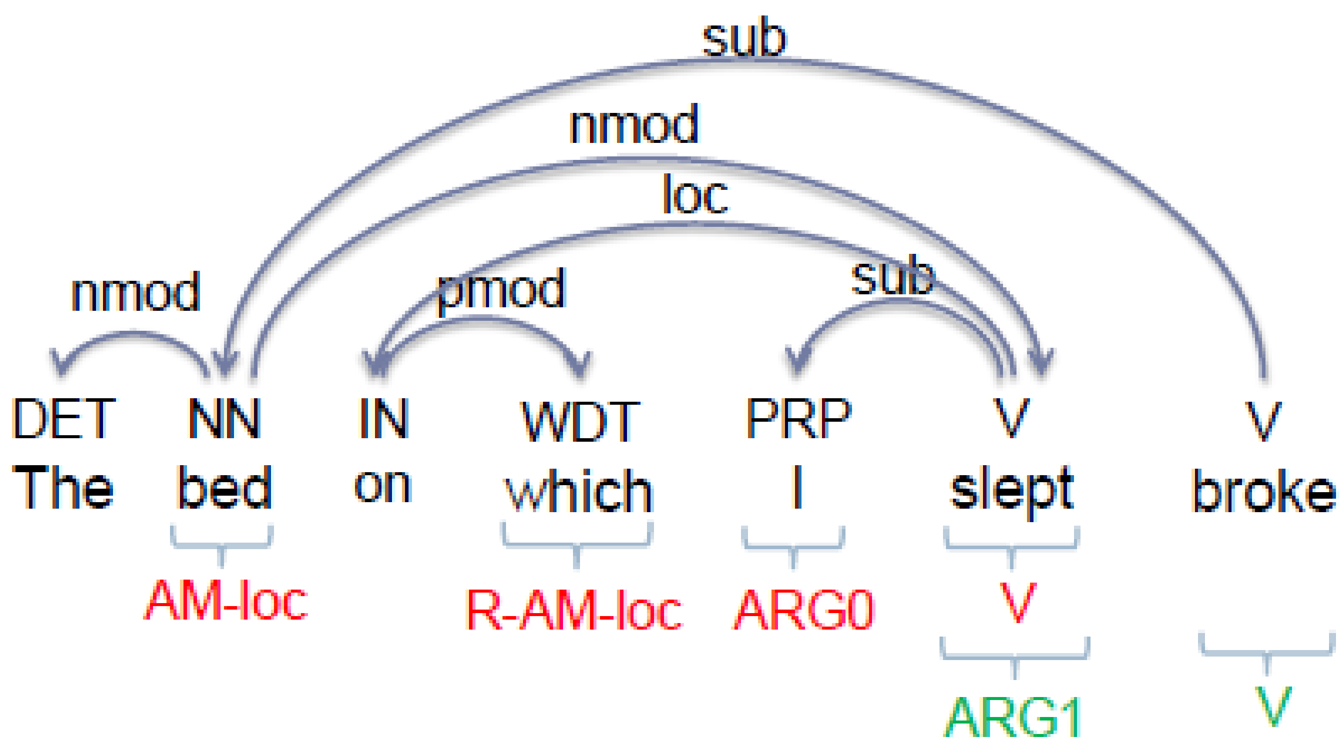
Semantic Role Labeling

SRL from constituent trees



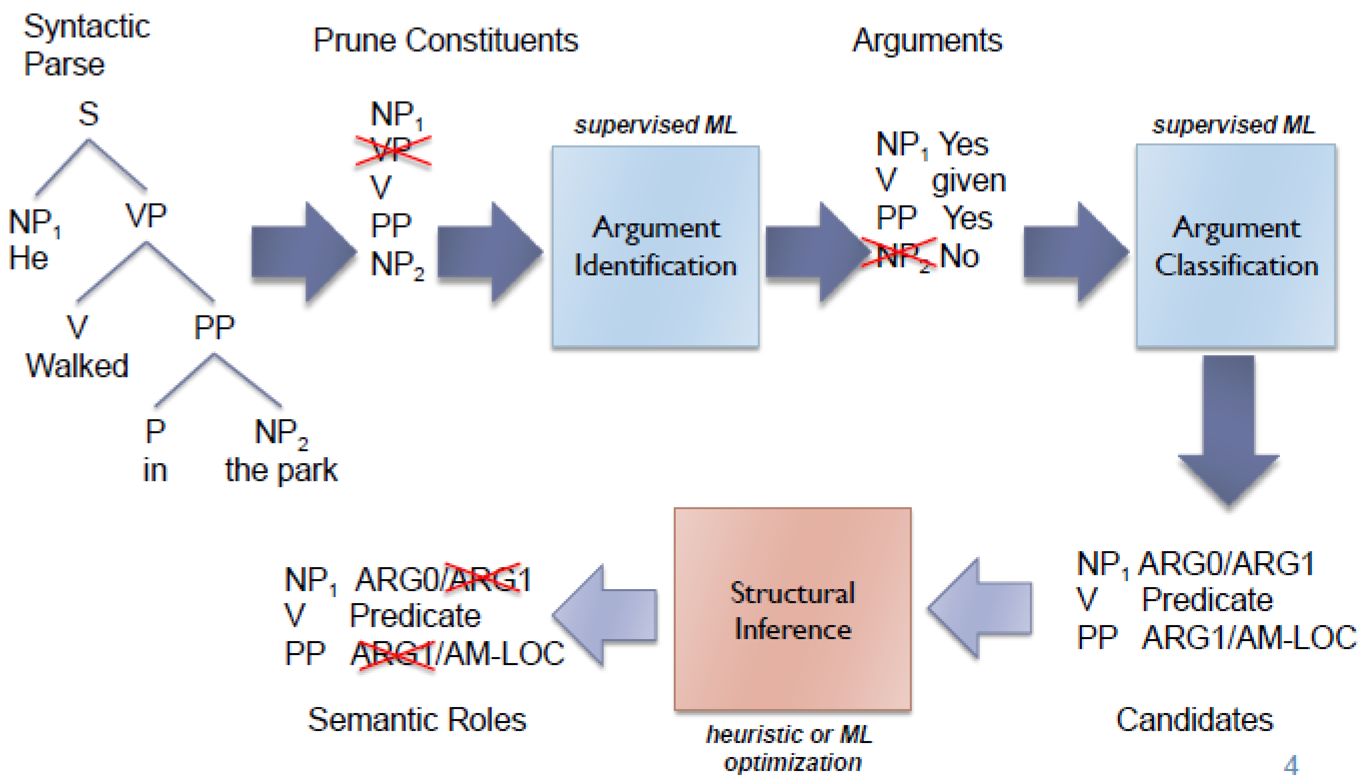
Semantic Role Labeling

SRL from dependency trees



Semantic Role Labeling

SRL supervised ML pipeline



Semantic Role Labeling

- Commonly used features:
  - Phrase types
    - verbal vs nominal realized predicates
    - explicit vs implicit args
  - Governing categories
  - Parse tree paths
  - Positions, distances
  - Voice
  - Tree kernels
  - Headword
  - Nes
  - Verb clusters
  - Previous role
  - Arg order

Semantic Role Labeling

- Commonly used ML models:
  - LibLinear
    - Perceptron
    - SVM
      - Linear and Tree Kernels
  - MaxEnt
  - Statistical Relational Models, SRM
  - Conditional Random Fields, CRF

Semantic Role Labeling

- Semi-supervised SRL (SSL) :
  - methods creating surrogate supervision: automatically annotate unlabeled data and treat it as new labeled data (annotation projection / bootstrapping methods)
  - parameter sharing methods: use unlabeled data to induce less sparse representations of words (clusters or distributed representations)
  - semi-unsupervised learning: adding labeled data (and other forms of supervision) to guide unsupervised models. Distant learning

Semantic Role Labeling

- Unsupervised SRL:
  - Goal: induce Semantic Roles automatically from unannotated texts
  - Approaches:
    - agglomerative clustering
      - Lang, Lapata, 2011
    - generative modelling
      - Titov, Klementiev, 2012
  - SRL is typically divided into two sub-tasks:
    - Identification: identification of predicate arguments
    - Labeling: assignment of their sematic roles

Semantic Role Labeling

- Agglomerative Clustering of argument keys:
  - Start with each argument key in its own cluster (high purity, low collocation)
  - Merge clusters together to improve collocation

- For a pair of clusters score:
  - whether a pair contains lexically similar arguments
  - whether arguments have similar parts of speech
  - whether the constraint that arguments in a clause should be in different roles is satisfied
    - *John taught students math*

- Prioritization
  - Instead of greedily choosing the highest scoring pair at each step, start with larger clusters and select best match for each of them

Semantic Role Labeling

- Generative modelling
  - Titov, Klementiev, 2012
- Bayesian Model

f or each predicate $p = 1, 2, \cdots$:
f or each occurrence l of p :
f or every role r *2* B_p :
if $[n \leftarrow Unif(0, 1)] = 1$:
GenArgument(p, r)
while $[n \leftarrow \;_{p,r}] = 1$:
    GenArgument(p, r)

GenArgument(p, r)

$k_{p,r} \leftarrow Unif(1, \ldots, |r|)$
$x_{p,r} \leftarrow ✓_{p,r}$

Meaning representation

- FOL
  - First Order Logic
- DRT
  - Discourse Representation Theory
- DL
  - Description Logic
- OWL
- others
  - ...

MR based on Logics

- Tutorials
  - Johan Bos
  - Alex Lascarides & Ewan Klein
  - David Ahn

MR based on Logics

DRT

- Shortcoming of FOL approaches to semantics
  - Anaphora across sentence boundaries
    - Pronouns:
      - John owns a car. It is red.
      - wrong: $\exists x(CAR(x) \wedge OWN(j, x)) \wedge RED(y)$
      - complex construction: $\exists x(CAR(x) \wedge OWN(j, x) \wedge RED(x))$
    - Problems with:
      - John doesn't own a car. ??It is red.
      - $\neg\exists x(CAR(x) \wedge OWN(j, x) \wedge RED(x))$
- Changing the Approach: Discourse Representation Theory
  - A new way of constructing LF
  - A new way of interpreting LF

MR based on Logics

DRT

- Text: *Vincent loves Mia*.
- DRT:

x y
vincent(x) mia(y) love(x,y)

- FOL: $\exists x \exists y(\text{vincent}(x)\ \&\ \text{mia}(y)\ \&\ \text{love}(x,y))$
- BK: $\forall x\ (\text{vincent}(x) \rightarrow \text{man}(x))$
  $\forall x\ (\text{mia}(x) \rightarrow \text{woman}(x))$
  $\forall x\ (\text{man}(x) \rightarrow \neg\ \text{woman}(x))$
- Model: $D = \{d1,d2\}$ $F(\text{vincent})=\{d1\}$
  $F(\text{mia})=\{d2\}$
  $F(\text{love})=\{(d1,d2)\}$

MR based on Logics

DRT

- If $x_1,\ldots, x_n$ are discourse referents and $\gamma_1,\ldots, \gamma_n$ are conditions, then

$$\begin{array}{|c|}\hline x_1 \;\ldots\; x_n \\ \hline \gamma_1 \;\ldots\; \gamma_n \\ \hline \end{array}$$

  is a DRS.
- If R is an n-ary relation symbol and $x_1,\ldots, x_n$ are discourse referents, then R($x_1,\ldots, x_n$) is a condition.
- If t_1 and t_2 are discourse referents, then $t_1 = t_2$ is a condition.
- If K_1 and K_2 are DRSs, then $K_1 \Rightarrow K_2$ is a condition.
- If K_1 and K_2 are DRSs, then $K_1 \vee K_2$ is a condition.
- If K is a DRS, then ¬K is a condition.

MR based on Logics

DRT

- The following DRS should be satisfied iff discourse referents x and y can be embedded (i.e., associated with entities in the model) such that:

$$\begin{array}{|c|}\hline x \quad y \\ \hline \texttt{woman}(x) \\ \texttt{boxer}(y) \\ \texttt{admire}(x, y) \\ \hline \end{array}$$

1. the first entity is a woman
2. the second is a boxer
3. the first stands in the *admires* relation to the second

MR based on Logics

DRT

- The discourse referents of a DRS K_1 are accessible from K_2 when:
  - K_1 **subordinates** K_2, or
  - K_1 **equals** K_2

MR based on Logics

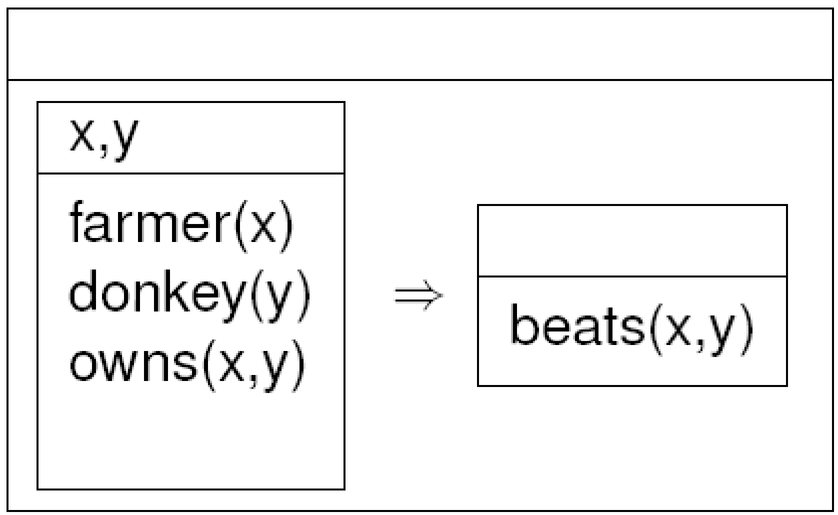
DRT

- DRS K_1 subordinates K_2 if and only if:
  - K_1 contains a condition of the form $\neg K_2$
  - K_1 contains a condition of the form $K_2 \Rightarrow K$, where K is some DRS
  - $K_1 \Rightarrow K_2$ is a condition in some DRS K
  - K_1 contains a condition of the form $K_2 \vee K$ or $K \vee K_2$, where K is some DRS
  - K_1 subordinates some DRS K, and K subordinates K_2
- In short: look up, and with implication, look left.

MR based on Logics

DRT

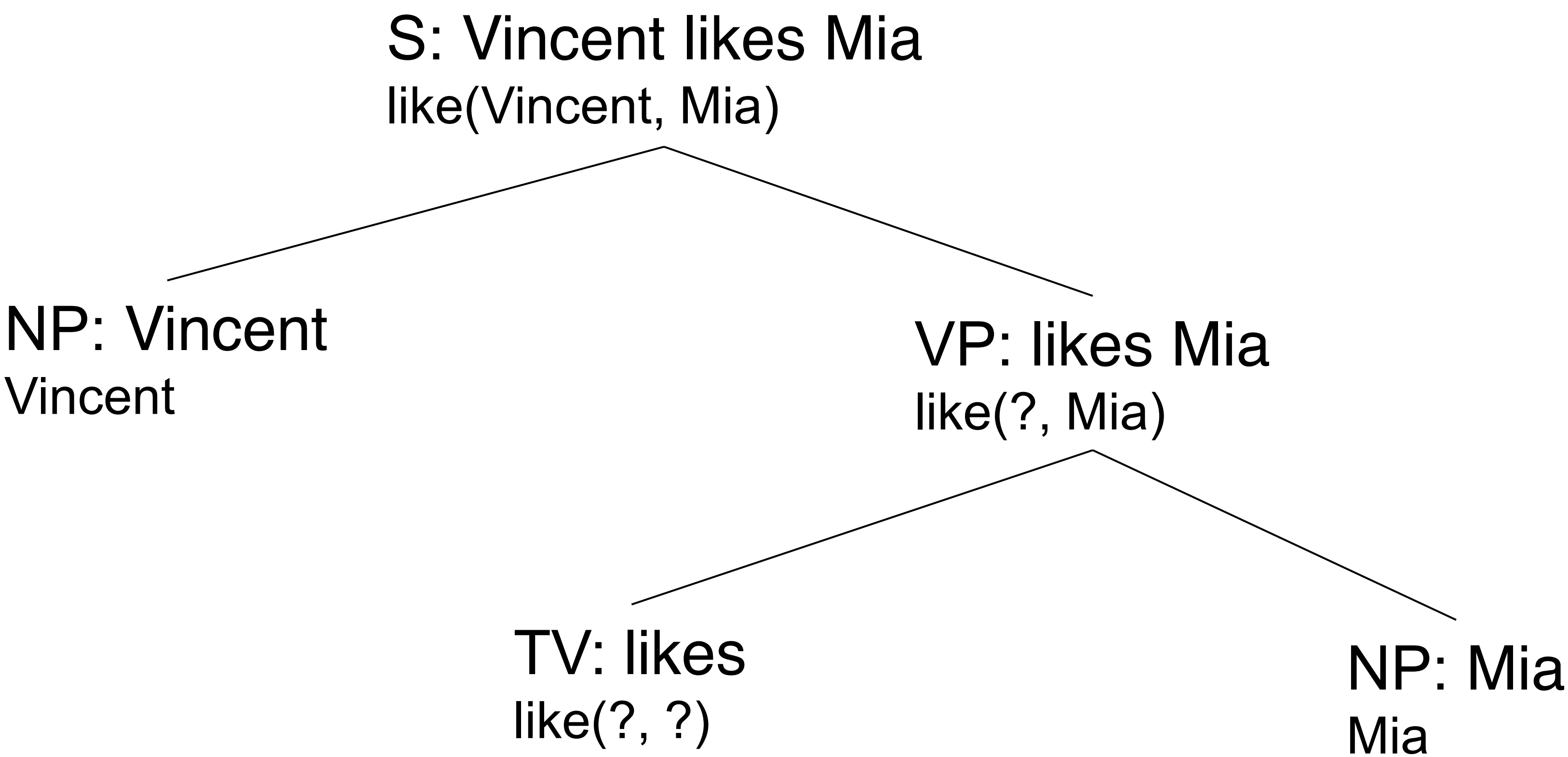
- *Every farmer who owns a donkey beats it*



MR based on Logics

Semantic Interpretation

Syntactic structure guides semantic construction



MR based on Logics

Semantic Interpretation

lambda calculus

$$\underbrace{\lambda x.\mathrm{man}(x)}_{\text{functor}} \ \underset{\text{functional application}}{@} \ \underbrace{\mathrm{vincent}}_{\text{argument}}$$

Fill each placeholder in the functor by an occurrence of the argument

MR based on Logics

Semantic Interpretation

lambda calculus

$\lambda x.man(x)$ @ vincent

β-conversion produces

man(vincent)

MR based on Logics

Semantic Interpretation

example: every boxer growls

- Step 1
  - Assign lambda expressions to the basic lexical items:
    - *boxer*: $\lambda y.boxer(y)$
    - *growls*: $\lambda x.growl(x)$
    - *every*: $\lambda P.\lambda Q.\forall x.(P(x) \rightarrow Q(x))$

MR based on Logics

Semantic Interpretation

example: every boxer growls

- Step 2
  - Associate the NP node with the application of the DET representation (functor) to the NOUN representation (argument)

NP: every boxer

$\lambda P.\lambda Q.\forall x(P@x \rightarrow Q@x)@\lambda y.boxer(y)$

DET: every

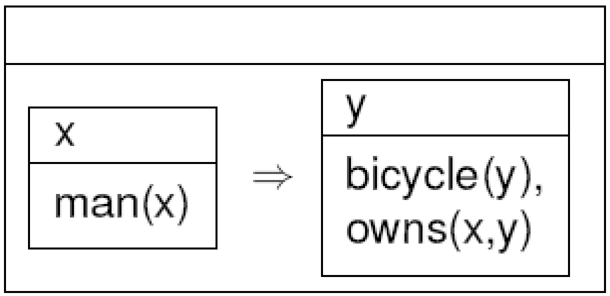
$\lambda P.\lambda Q.\forall x(P@x \rightarrow Q@x)$

Noun: boxer

$\lambda y.boxer(y)$

MR based on Logics

Semantic Interpretation in DRT



DRS in NLTK

DRS([],[(DRS([x],[(man x)]) implies DRS([y],[(bicycle y),(owns y x)]))])

toFol(): Converts DRSs to FoL.

draw(): Draws a DRS in ‘box’ notation

MR based on Logics

DRT to FOL

- In order to use first-order inference tools to verify acceptability constraints, we need to translate DRT into FOL (w/equality).
- Translation is performed by translation function *t*.
- $(arg)^t$ indicates the application of *t* to *arg* (i.e., the translation of *arg*), where *arg* is either a DRS or a condition.

MR based on Logics

DRT to FOL

$$\left(\begin{array}{|c|}\hline x_1 \ \ldots \ x_n \\ \hline \gamma_1 \\ \cdot \\ \cdot \\ \cdot \\ \gamma_m \\ \hline\end{array}\right)^t = \exists x_1 \cdots \exists x_n.((\gamma_1)^t \wedge \cdots \wedge (\gamma_m)^t)$$

MR based on Logics

DRT to FOL

$$\left(\begin{array}{|c|}\hline \\ \hline \gamma_1 \\ \cdot \\ \cdot \\ \cdot \\ \gamma_m \\ \hline\end{array}\right)^t = (\gamma_1)^t \wedge \cdots \wedge (\gamma_m)^t$$

MR based on Logics

DRT to FOL

- $(\mathrm{R}(x_1, \ldots, x_n))^t = \mathrm{R}(x_1, \ldots, x_n)$
- $(\ x_1 = x_2\ )^t = x_1 = x_2$
- $(\ x = c\ ) = x = c$
- $(\ c = x\ ) = c = x$
- $(\neg \mathrm{K})^t = \neg(\mathrm{K})^t$
- $(\mathrm{K}_1 \vee \mathrm{K}_2)^t = (\mathrm{K}_1)^t \vee (\mathrm{K}_2)^t$

MR based on Logics

DRT to FOL

$$\left(\begin{array}{|c|} \hline x_1 \ \ldots \ x_n \\ \hline \gamma_1 \\ \cdot \\ \cdot \\ \cdot \\ \gamma_m \\ \hline \end{array} \Rightarrow K \right)^t = \forall x_1 \cdots \forall x_n . (((\gamma_1)^t \wedge \cdots \wedge (\gamma_m)^t) \rightarrow K^t)$$

MR based on Logics

DRT to FOL

$$\left(\begin{array}{|c|} \hline \\ \hline \gamma_1 \\ \cdot \\ \cdot \\ \cdot \\ \gamma_m \\ \hline \end{array} \Rightarrow K \right)^t = ((\gamma_1)^t \wedge \cdots \wedge (\gamma_m)^t) \to K^t$$

Semantic Parsing

Wide-coverage semantic parsers

- **Lingo/LKB**
  - Minimal Recursive Semantics
  - [Copestake 2002]
- **Shalmaneser**
  - Frame Semantics
  - [Erk & Pado 2006]
- **Boxer**
  - Discourse Representation Structures
  - [Bos 2005]

Semantic Parsing

Boxer

- **Lexical Semantics**
  - Lambda calculus as glue language
  - Function application and beta-conversion
- **Semantic formalism**
  - DRS
  - FOL
- **Output format**
  - Prolog terms
  - XML

Semantic Parsing

C&C

- C&C tools
  - CCG Parser
    - CCGbank
      - treebank of CCG derivations developed by Julia Hockenmaier and Mark Steedman
      - semi-automatically converting the phrase-structure trees in the Penn Treebank
    - Parser & Grammar
      - Wide-Coverage Efficient Statistical Parsing with CCG and Log-Linear Models
      - http://web.comlab.ox.ac.uk/oucl/work/stephen.clark/papers/cl07parser.pdf
  - Boxer
    - James R. Curran, Stephen Clark, and Johan Bos (2007)

Semantic Parsing

C&C

- CCG Parser
  - Lexical Category Set
    - 425 different categories
  - Combinatory Rules
    - Combine categories
      - forward and backward application
      - forward composition
      - generalised forward composition
      - backward composition
      - backward-crossed composition
      - generalised backward-crossed composition
      - type-raising
      - coordination schema which coordinates any two categories of the same type

Semantic Parsing

C&C

- Boxer
  - produces standard DRS syntax
  - uses a neo-Davidsonian analysis for events
  - incorporates Van der Sandt's algorithm for presupposition
  - is 100% compatible with first-order logic (FOL)
  - normalises cardinal and date expressions
  - DRSs can be generated in various output formats:
    - resolved or underspecified,
    - in Prolog or XML,
    - flattened or recursive structures,
    - with discourse referents represented by Prolog atoms or variables,
    - with pretty printed DRSs or not.
    - It is also possible to output FOL formulas translated from the DRSs

Semantic Parsing

C&C

- Example
  - Every man runs
- parsing (CCG)
  - ```
    ccg(1,
     rp('S[dcl]',
      ba('S[dcl]',
       fa('NP[nb]',
        lf(1,1,'NP[nb]/N'),
        lf(1,2,'N')),
       lf(1,3,'S[dcl]\NP')),
      lf(1,4,'.'))).
    ```
  - ```
    w(1, 1, 'Every', 'every', 'DT', 'I-NP', 'O', 'NP[nb]/N').
    w(1, 2, 'man', 'man', 'NN', 'I-NP', 'O', 'N').
    w(1, 3, 'runs', 'run', 'VBZ', 'I-VP', 'O', 'S[dcl]\NP').
    w(1, 4, '.', '.', '.', 'O', 'O', '.').
    ```

Semantic Parsing

C&C

- Semantic analysis (Boxer)
  - ```
    sem(1,
         [
          word(1001, 'Every'),
          word(1002, man),
          word(1003, runs),
          word(1004, '.')
         ],
         [
          pos(1001, 'DT'),
          pos(1002, 'NN'),
          pos(1003, 'VBZ'),
          pos(1004, '.')
         ],
         [
         ],...
    ```
    ```

Semantic Parsing

C&C

- Semantic analysis (Boxer)
  - ```
    %%%  __________________________________________
    %%% |                                          |
    %%% |__________________________________________|
    %%% |   __________          ________________   |
    %%% |  | x1       |        | x2             |  |
    %%% |  |__________|        |________________|  |
    %%% |  | man(x1)  |  ==>   | run(x2)        |  |
    %%% |  |__________|        | event(x2)      |  |
    %%% |                      | agent(x2,x1)   |  |
    %%% |                      |________________|  |
    %%% |__________________________________________|
    ```
    ```

Description Logic

Description Logic (DL)

- (from wikipedia)
- Family of knowledge representation languages which can be used to represent the terminological knowledge of an application domain
- Extension to frames and semantic networks, which were not equipped with formal logic-based semantics.
- KL-ONE (1985), LOOM (1987), ... RACER (2001), KAON 2 (2005).

Description Logic

Description Logic (DL)

- Modelling in Description Logics.
  - TBox (terminological box)
    - In general, the TBox contains sentences describing concept hierarchies (i.e., relations between concepts)
  - ABox (assertional box).
    - The ABox contains "ground" sentences stating where in the hierarchy individuals belong (i.e., relations between individuals and concepts).
  - Example
    - (1) Every employee is a person
    - belongs in the TBox
    - (2) Bob is an employee
    - belongs in the ABox

Description Logic

Description Logic (DL)

- DL Reasoners.
  - Pellet, an open-source Java OWL DL reasoner
  - FaCT, a DL classifier
  - FaCT++, the new generation of FaCT OWL-DL reasoner
  - KAON2 is a free (free for non-commercial usage) Java reasoner
  - RacerPro is a commercial (free trials and research licenses are available) lisp-based reasoner.
- Other tools
  - Protégé is a free, open source ontology editor and knowledge-base framework, which can use DL reasoners which offer a DIG interface as backends for consistency checks.
  - DIG Implementation. DIG is an XML interface to DL systems
  - SPARQL Query Language for RDF

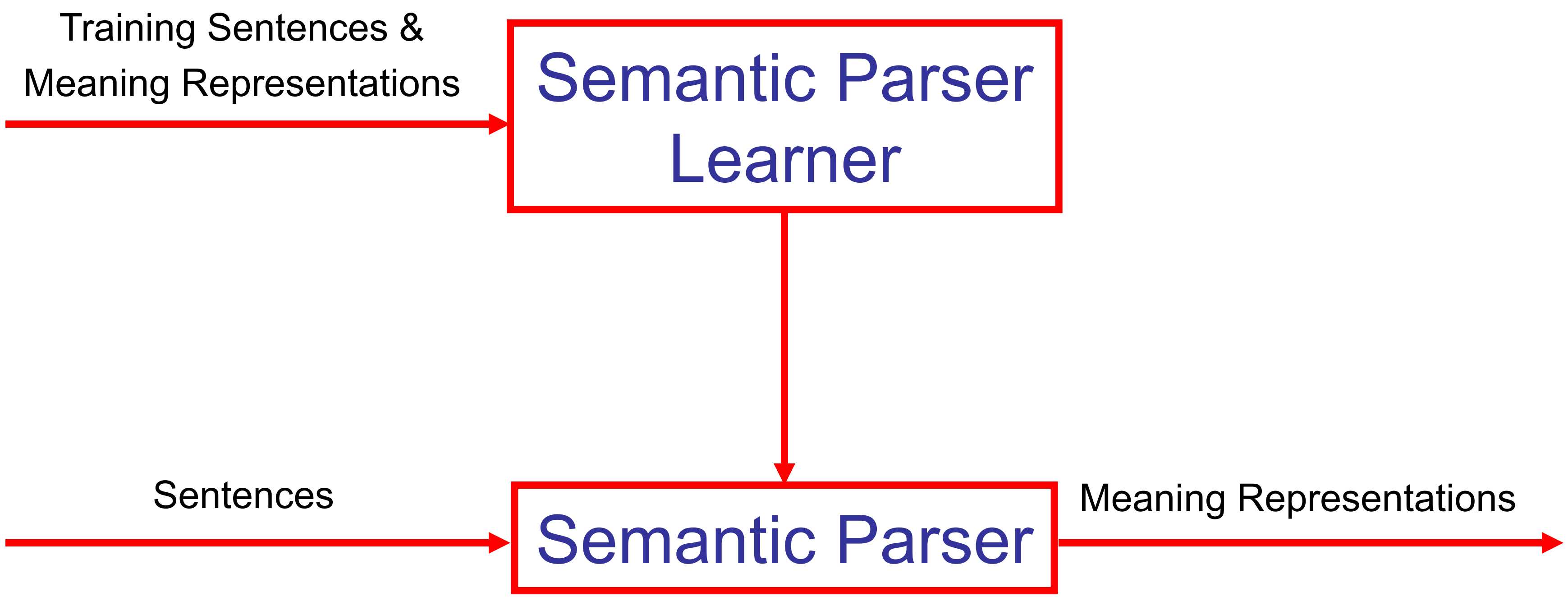
Learning Semantic Parsers

- Supervised approaches on narrow domains
- Semi-supervised approaches
  - Distant Learning
  - Indirect Learning
- Unsupervised approaches

Learning Semantic Parsers

- Supervised approaches on narrow domains
- Seminal Work at Texas University (Raymond Mooney)
- Thesis at TU
  - Rohit J. Kate (2007)
  - Yuk Wha Wong (2007)
  - Ruifang Ge (2010)
  - David L. Chen (2012)
  - Joohyun Kim (2013)
- ACL 2010 Tutorial
  - Rohit J. Kate & Yuk Wah Wong

Learning Semantic Parsers



Learning Semantic Parsers

- Transforming a natural language sentence into its meaning representation
- Example application domains (very narrow)
  - ATIS: Air Travel Information Service
  - CLang: Robocup Coach Language
  - Geoquery: A Database Query Application
  - Virtual worlds from the navigation tasks

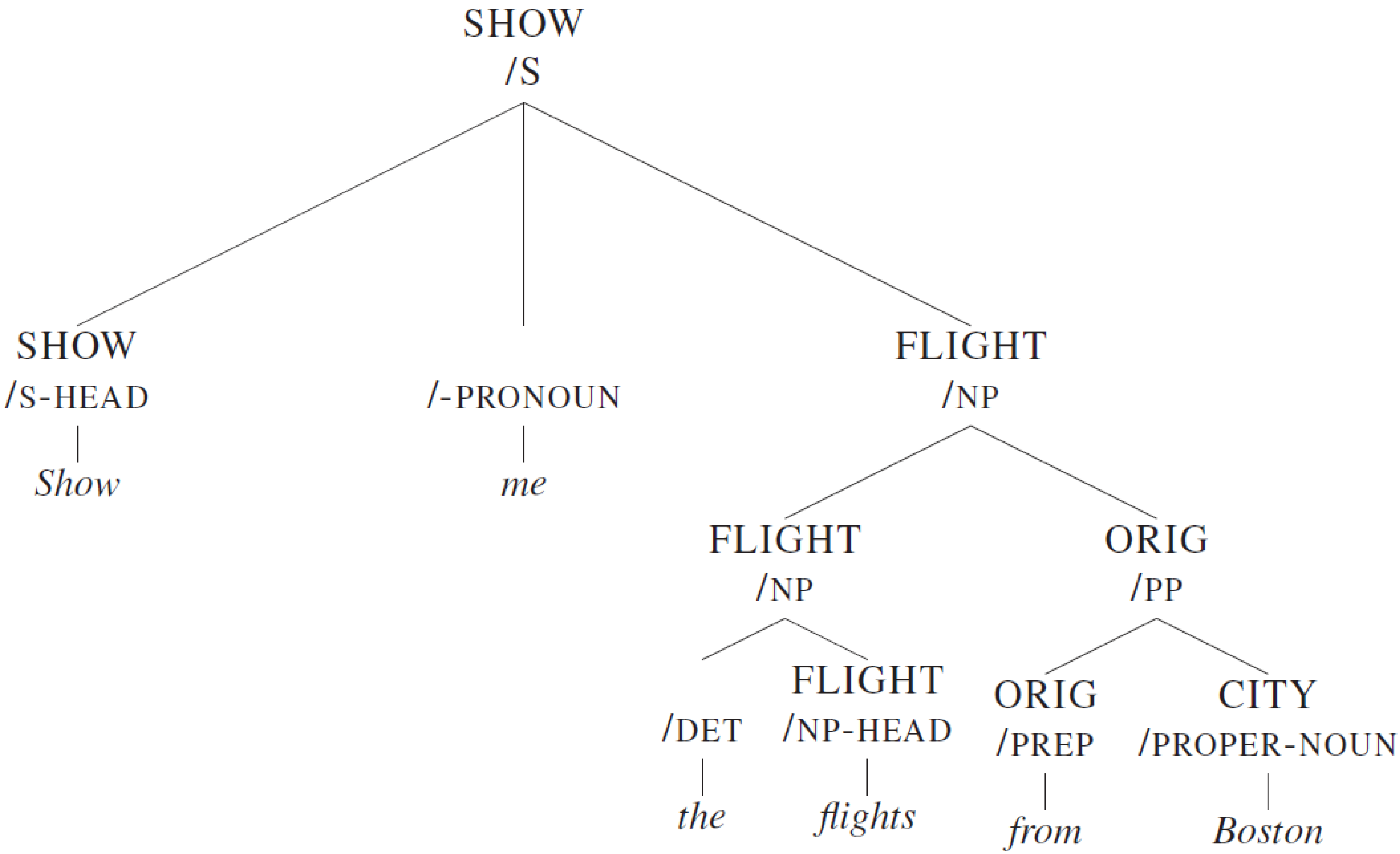
Robocup Coach Language

If our player 2 has the ball, then position our player 5 in the midfield.

```
((bowner (player our {2}))
 (do (player our {5}) (pos (midfield))))
```

300 pieces of coaching advice
22.52 words per sentence

ATIS corpus


SHOW
/S
SHOW
/S-HEAD
Show
/-PRONOUN
me
FLIGHT
/NP
FLIGHT
/NP
/DET
the
FLIGHT
/NP-HEAD
flights
ORIG
/PP
ORIG
/PREP
from
CITY
/PROPER-NOUN
Boston

Geoquery

880 queries on a geography database
7.48 word per sentence
MRL: Prolog and FunQL

What are the rivers in Texas?

answer(x_1,(river(x_1),loc(x_1,x_2),equal(x_2,stateid(texas))))

Learning Semantic Parsers

- Initial system
  - Inductive logic programming (Zelle & Mooney, 1996)
- Current approaches
  - Tang & Mooney, 2001
    - COCKTAIL
    - Deterministic, inductive logic programming
  - Zettlemoyer & Collins (2005, 2007)
    - Structured learning with combinatory categorial grammars (CCG)
  - Wong & Mooney (2006, 2007a, 2007b)
    - Syntax-based machine translation methods
  - Kate & Mooney (2006), Kate (2008a)
    - SVM with kernels for robust semantic parsing
  - Lu et al. (2008)
    - A generative model for semantic parsing
  - Ge & Mooney (2005, 2009)
    - Exploiting syntax for semantic parsinG

WASP

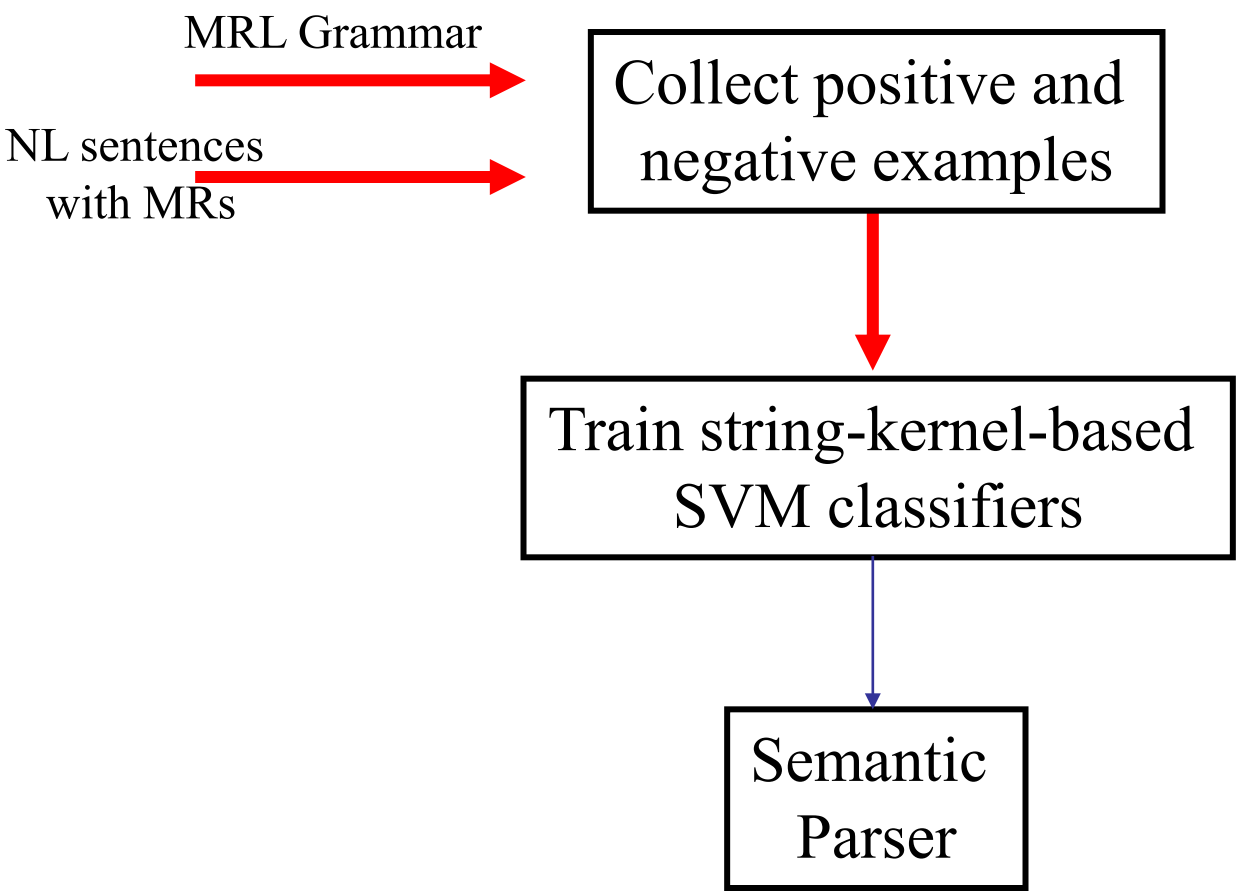
- A Machine Translation Approach to Semantic Parsing
  – Wong & Mooney (2006)
- Based on a semantic grammar of the natural language
- Uses machine translation techniques
  – Synchronous context-free grammars
  – Word alignments

KRISP

- **K**ernel-based **R**obust **I**nterpretation for **S**emantic **P**arsing
  - Kate & Mooney (2006), Kate (2008)
- Learns semantic parser from NL sentences paired with their respective MRs given MRL grammar
- Productions of MRL are treated like semantic concepts
- A string classifier is trained for each production to estimate the probability of an NL string representing its semantic concept
- These classifiers are used to compositionally build MRs of the sentences

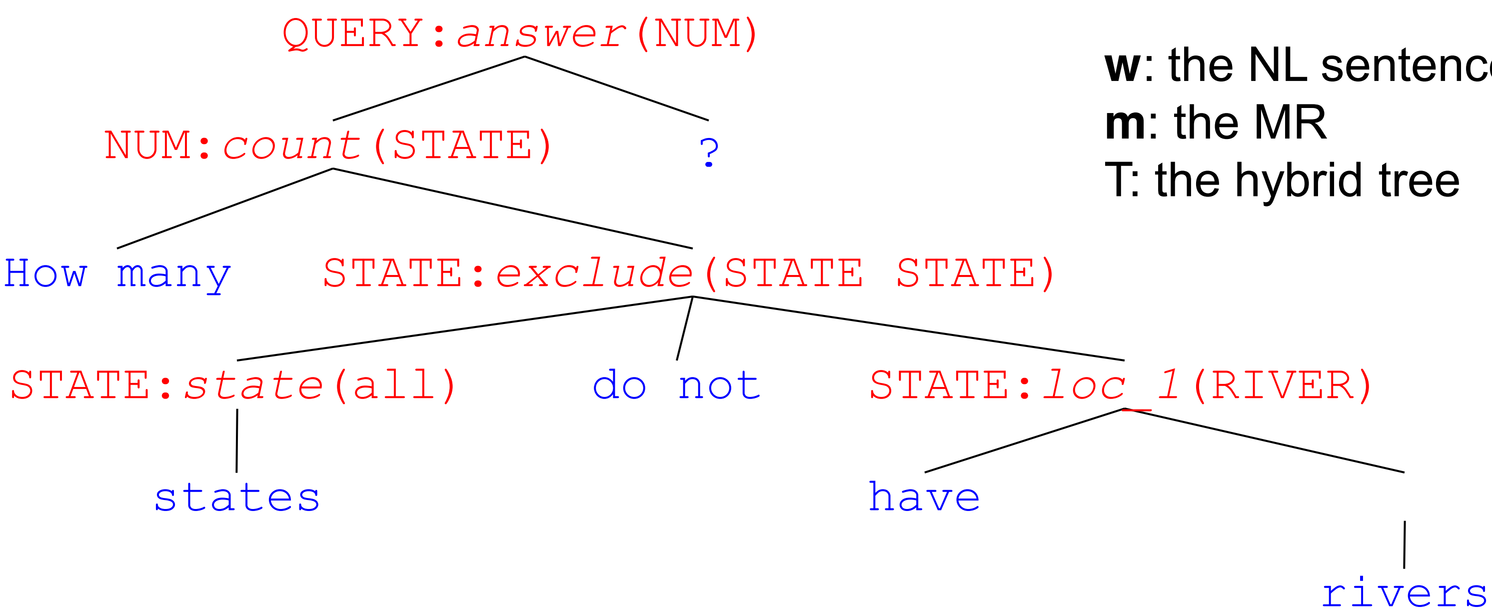
Overview of KRISP

Training



A Generative Model

- A Generative Model for Semantic Parsing
- Hybrid Tree
- Lu et al, 2008

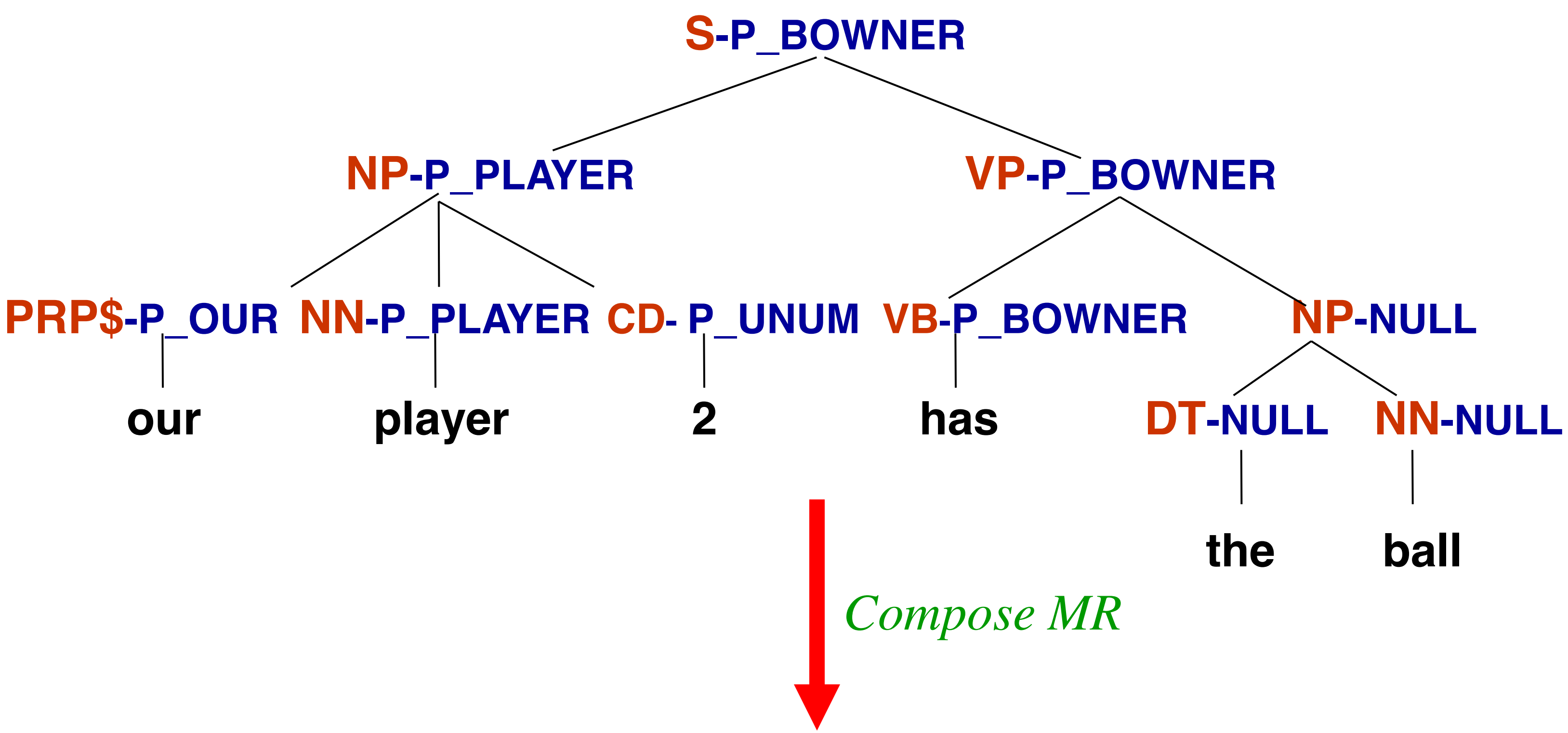


w: the NL sentence
m: the MR
T: the hybrid tree

SCISSOR

- Ge & Mooney (2005)
- ***S***emantic ***C***omposition that ***I***ntegrates ***S***yntax and ***S***emantics to get ***O***ptimal ***R***epresentations
- Integrated syntactic-semantic parsing
  – Allows both **syntax and semantics** to be used **simultaneously** to obtain an accurate combined syntactic-semantic analysis
- A statistical parser is used to generate a semantically augmented parse tree (SAPT)

SAPT



MR: (bowner (player our {2}))

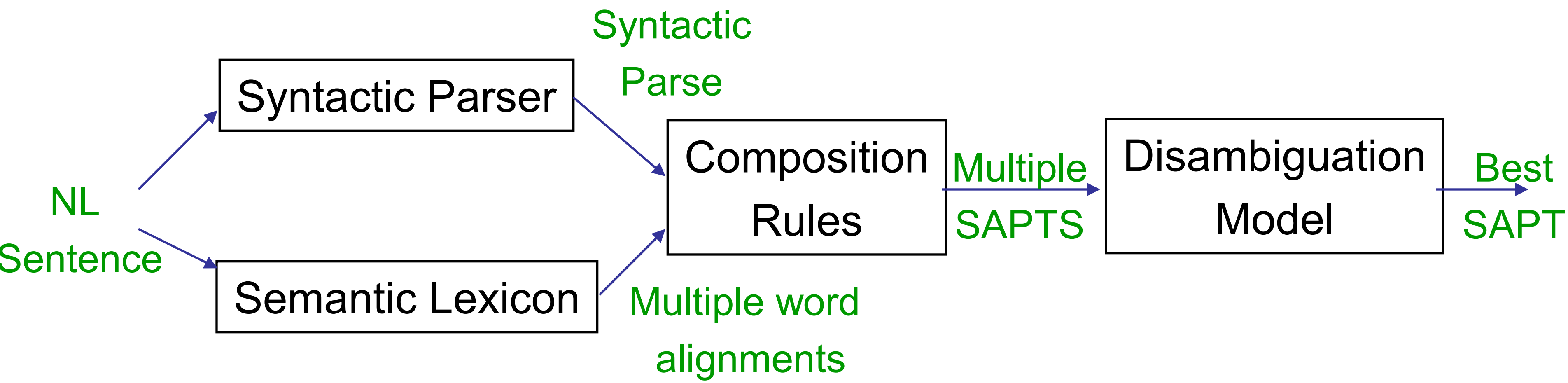
Results on CLang

	Precision	Recall	F-measure
SCISSOR	89.5	73.7	80.8
WASP	88.9	61.9	73.0
KRISP	85.2	61.9	71.7
LU	82.4	57.7	67.8

SYNSEM

- Ge & Mooney (2009)
- **SCISSOR** requires extra SAPT annotation for training
- Must learn both syntax and semantics from same limited training corpus

SYNSEM Overview Ge & Mooney (2009)



KRISPER

- KRISP with **E**M-like **R**etraining
- Kate & Mooney 2007
- Extension of KRISP that learns from *ambiguous* supervision
- Uses an iterative Expectation-Maximization-like method to gradually converge on a correct meaning for each sentence
- Successfully learns semantic parser with ambiguous supervision

EMBEDDED SYSTEMS

- Embedding Methods for NLP
  - Weston & Bordes, EMNLP tutorial 2014
- Deep Learning
- Similar words should have similar embeddings (share latent features).
- Embeddings can also be applied to symbols as well as words (e.g. Freebase nodes and edges).
- Can also have embeddings of phrases, sentences, documents, or even other modalities such as images.

EMBEDDED SYSTEMS

- Embedding Models
  - Models based on low-dimensional continuous vector embeddings for entities and relation types, directly trained to define a similarity criterion.
  - Stochastic training based on ranking loss with sub-sampling of unknown relations.

EMBEDDED SYSTEMS

- Latent semantic indexing (LSI)
  - Learn a linear embedding
- Neural Net Language Models (NN-LMs) (Bengio et al., ’06)
- Recurrent NN-LMs (Mikolov et al., ’10).
- SENNA, (Collobert, Weston, 2008)
- Wsabie, (Weston et al 2010)
- Word2Vec (Mikolov et al., ’13).
- RNN, (Socher et al, 2011)
- Neural Tensor Networks, (Socher et al, 2013)

EMBEDDED SYSTEMS

- Embedding Models for KBs
- Subjects and objects are represented by vectors in the embedding space.
- Rel. types = similarity operators between subj/obj.
- Learning similarities depending on
  - rel -> <sub,rel,obj)
  - parameterized by **s**, **R** and **o**.

EMBEDDED SYSTEMS

- Modeling Relations as Translations
  - (Bordes et al, 2013)
  - **s + r ≈ o**

- Subgraph Embeddings (Bordes et al., ’14)
- Model learns embeddings of questions and (candidate) answers
- Answers are represented by entity and its neighboring subgraph

EMBEDDED SYSTEMS

- Code
  - Torch: www.torch.ch
  - SENNA: ronan.collobert.com/senna
  - RNNLM: www.fit.vutbr.cz/~imikolov/rnnlm
  - Word2vec: code.google.com/p/word2vec
  - Recursive NN: nlp.stanford.edu/sentiment
  - SME (multi-relational data): github.com/glorotxa/sme

MRD

- Multi-relational data
  - Data is structured as a graph
  - Each node = an entity
  - Each edge = a relation/fact
  - A relation = (sub, rel , obj):
    - sub =subject,
    - rel = relation type,
    - obj = object.
  - Nodes w/o features.

MRD

- Scaling semantic parsers to large knowledge bases has attracted substantial attention recently
  - Cai and Yates, 2013
  - Berant et al. 2013
  - Kwiatkowski et al., 2013

Q&A over LD

- Open-domain Question Answering
- answer question on any topic
  - query a KB with natural language
  - Semantic Representation = KB entities + relations

Q&A over LD

- Question Answering with Subgraph Embeddings
  - A. Bordes, S. Chopra & J. Weston. EMNLP, 2014
- Paraphrase-Driven Learning for Open Question Answering
  - A. Fader, L. Zettlemoyer & O. Etzioni. ACL, 2013
- Open Question Answering Over Curated and Extracted Knowledge Bases
  - A. Fader, L. Zettlemoyer & O. Etzioni. KDD, 2014
- Large-scale Semantic Parsing without Question-Answer Pairs
  - S. Reddy, M. Lapata & M. Steedman. TACL, 2014.

Q&A over LD

- Question Answering with Subgraph Embeddings
  - Training data
    - Freebase is automatically converted into Q&A pairs closer to expected language structure than triples

QALD contests

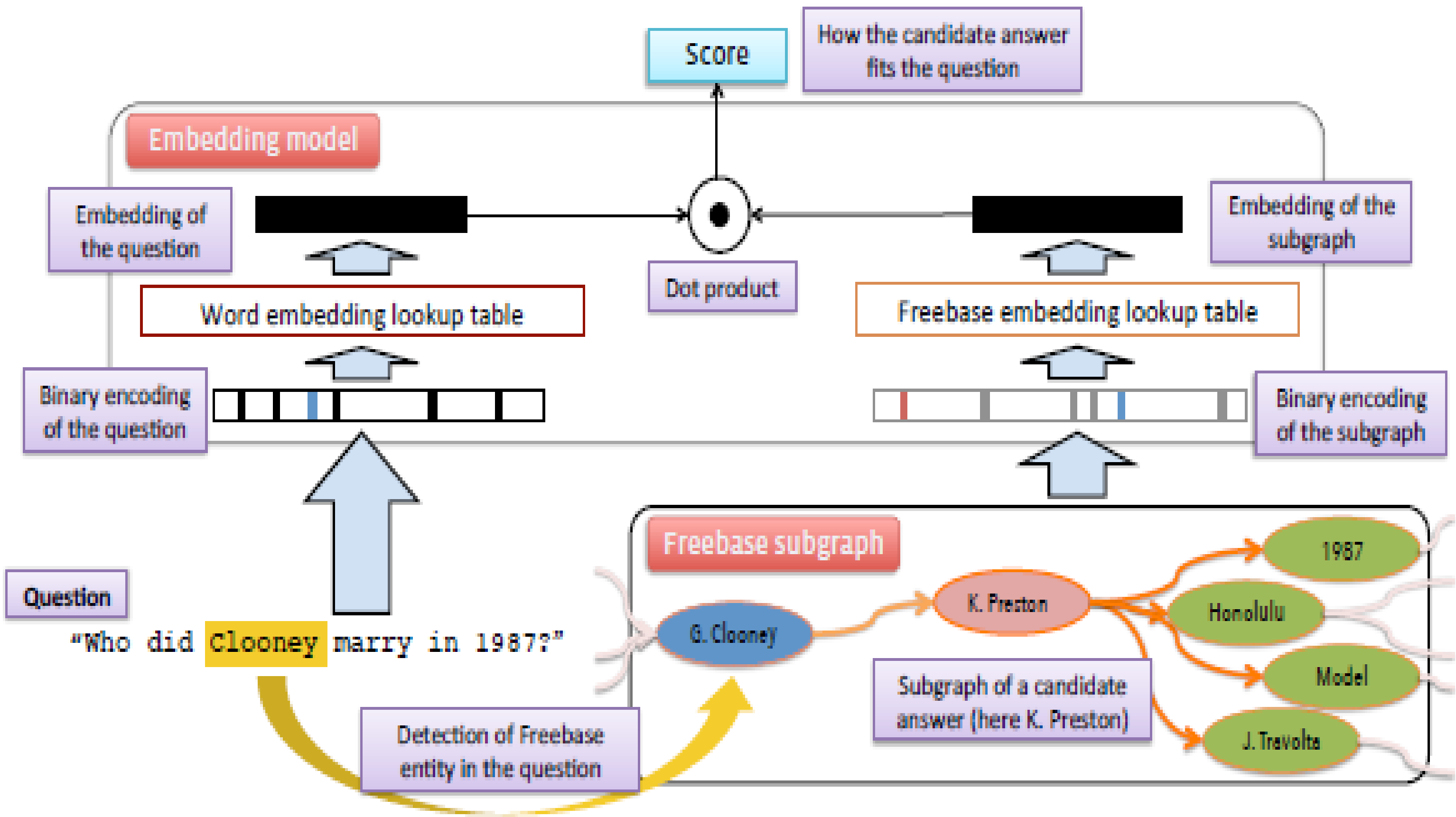
- QALD5
  - http://greententacle.techfak.uni-bielefeld.de/~cunger/qald/index.php?x=challenge&q=5
  - Given a natural language question or keywords, retrieve the correct answer(s) from a repository containing both RDF data and free text.
  - QALD1, …, QALD5

QALD contests

- < question id ="272"
- answertype =" resource “
- aggregation =" true “
- onlydbo =" true “
- hybrid =" false " >
- < string lang =" en " > Which book has the most pages
- < string lang =" es " >¿Que libro tiene el mayor numero de paginas ?
- < keywords lang =" en " > book , the most pages
- < query >
  - PREFIX dbo : < http :// dbpedia . org / ontology / >
  - PREFIX rdf : < http :// www . w3 . org /1999/02/22 - rdf - syntax - ns # >
  - SELECT DISTINCT ?uri
  - WHERE {
    - ?uri rdf : type dbo : Book .
    - ?uri dbo : numberOfPages ? n }
  - ORDER BY DESC (?n) OFFSET 0 LIMIT 1

Q&A over LD

Bordes et al, 2014



Q&A over LD

- Fader (2013) in his thesis presented a QA system that maps questions onto simple queries against Open IE extractions, by learning paraphrases from a large monolingual parallel corpus, and performing a single paraphrasing step.
  - PARASEMPRE
    - http://www-nlp.stanford.edu/software/sempre/.

Q&A over LD

- PARALEX
  - large monolingual parallel corpora, containing 18 million pairs of question paraphrases from wikianswers.com, which were tagged as having the same meaning by users.
- PARALEX focuses on question paraphrases.

Q&A over LD

Fader thesis, 201